STUDIES ON REGULARIZED NEWTON-TYPE METHODS FOR UNCONSTRAINED MINIMIZATION PROBLEMS AND THEIR GLOBAL COMPLEXITY BOUNDS

KENJI UEDA

STUDIES ON REGULARIZED NEWTON-TYPE METHODS FOR UNCONSTRAINED MINIMIZATION PROBLEMS AND THEIR GLOBAL COMPLEXITY BOUNDS

by

KENJI UEDA

Submitted in partial fulfillment of
the requirement for the degree of
DOCTOR OF INFORMATICS
(Applied Mathematics and Physics)

KYOTO UNIVERSITY
KYOTO, JAPAN
JANUARY 2012

Preface

In this thesis, we consider the unconstrained minimization problem and its related problems, such as the system of equations, the nonlinear least squares problem and the nonlinear complementarity problem (NCP). These problems have various applications in the real world, and they are the fundamental problems in nonlinear optimization problems. Thus, it is worth studying on these problems not only for their own applications but also for constructing more general-purpose solution methods. Many solution methods for solving these problems, such as the steepest descent method and the Newton-type methods, have been proposed. However, there still remain a lot of problems that cannot be solved in practical time by the existing solution methods.

Efficiencies of solution methods for our problems are often discussed in terms of the global convergence or the rapid rate of convergence. The global convergence property ensures to get a solution from an arbitrary initial point. On the other hand, the rate of convergence represents the speed to get a solution from a point near the solution. Recently, in addition to these points of view, global complexity bounds of solution methods have received much attention, and have been actively investigated. The global complexity bound is defined as an upper bound of the number of iterations required to get a reasonable solution satisfying appropriate conditions. Since it corresponds to the worst computational time, it is useful when we want to estimate in advance the time for solving a large-scale problem.

The regularization technique is a basic technique in several fields. For example, by using the regularization in the convex optimization problem, an original convex problem can be reformulated to a strongly convex problem, which is much easier to deal with. The regularized Newton-type methods, such as the regularized Newton method (RNM) and the Levenberg-Marquardt method (LMM), are based on the Newton method, and exploit the regularization to get a search direction. For convex problems, many RNMs have been proposed and shown to have nice convergence properties. However, the RNM for nonconvex problems has not been well studied because it is more difficult to theoretically analyze its convergence properties. On the other hand, the regularization is directly exploited with the (Gauss-)Newton method for the nonlinear least squares problem or the system of equations. The LMM is a regularized Gauss-Newton method, and has global and superlinear convergence. However, its global complexity bound remains unknown. In this thesis, we study on the regularized Newton-type methods for the unconstrained minimization problem and its related problems, and on their global complexity bounds.

The first contribution of this thesis is to propose RNMs extended to the unconstrained nonconvex minimization problem, and to estimate their global complexity bounds. First, we develop an extended RNM using a certain line search, and investigate its convergence properties, in particular global and superlinear convergence and the global complexity bound. Second, we propose an RNM without line search. The method uses a technique similar to the trust-region method, and is shown to have global convergence. Moreover, we show that the proposed RNM have nice theoretical and numerical properties.

Another contribution is to estimate global complexity bounds of LMMs. First, we give the global complexity bound of the LMM for the nonlinear least squares problem with a smooth mapping. Second, we also give the same global complexity bound of the LMM for the system of nonsmooth equations. The system of nonsmooth equations includes important applications such as the NCP. Thus, the obtained global complexity bound helps us to get the bound for the NCP.

The author hopes that the results of this thesis will contribute for further studies on developing efficient solution methods for unconstrained minimization problems and their related problems.

Kenji Ueda
January 2012

Acknowledgment

First of all, I would like to express my sincere appreciation to Professor Masao Fukushima of Kyoto University. I thank him for accepting me to enter the doctoral course of Kyoto University, though I am a full-time worker. He gave me the invaluable opportunity, and this has been one of the greatest opportunities in my life. His eager and earnest attitude to the studies is respectable. I would like to bear such a spirit in my mind.

Second, I would also like to express my deepest gratitude to Associate Professor Nobuo Yamashita of Kyoto University, who supervised me throughout my doctoral course. Although I may have often troubled him, he kindly gave me so many constructive comments, and spared precious time for me to discuss. His sincere and caring attitude toward his students is so admirable and it makes me want to be a person like him. Without his continuous guidance and persistent support, it would not have been possible for me to write this thesis.

Since April 2009, I have worked at Mitsubishi Electric Corporation. Satoru Takahashi, Manager of Traffic Control Systems Group, was willing to accept my enrollment in the doctoral course. I would like to express my appreciation for his consideration.

I would also like to tender my acknowledgments to Assistant Professor Shunsuke Hayashi of Kyoto University and Assistant Professor Hiroshige Dan of Kansai University. They gave me a lot of advices when I faced my difficulties in balancing study with work. In addition, I would like to thank my colleagues of Mitsubishi Electric Corporation and all members of Fukushima Laboratory for their warm supports.

Finally, I would like to dedicate this thesis to my parents with my appreciation for their tireless supports and encouragements.

Contents

Notations

$\lambda_{\max}(M)$	the maximum eigenvalue of a symmetric matrix M
$\lambda_{\min}(M)$	the minimum eigenvalue of a symmetric matrix M
$B(x,r)$	the closed sphere with center x and radius r
$\mathrm{dist}(x,S)$	the distance between a vector x and a set S
ε_f	the accuracy on f used in the global complexity bound
ε_g	the accuracy on the norm of ∇f used in the global complexity bound
L_f	the Lipschitz constant of f
L_g	the Lipschitz constant of ∇f
L_H	the Lipschitz constant of $\nabla^2 f$
L_F	the Lipschitz constant of F
L_J	the Lipschitz constant of ∇F
$\partial F(x)$	the generalized Jacobian of F at x
Ω	the level set of f at the initial point x^0
x^*	the local optimal solution
X^*	the local optimal solution set
f_k	the value of f at x^k
g^k	the gradient of f at x^k
H_k	the Hessian of f at x^k
F^k	the point of F at x^k
J_k	the Jacobian or the element of the generalized Jacobian of F at x^k
d^k	the search direction at the k-th iteration
Λ_k	$\max(0, -\lambda_{\min}(H_k))$
l_k	the number of inner iteration
μ_k	the regularization parameter at the k-th iteration
ϕ_k	the model function of f at x^k
ρ_k	the ratio of the reduction of f to that of ϕ_k

Chapter 1

Introduction

1.1 Unconstrained minimization problem and its related problems

In this thesis, we consider the unconstrained minimization problem.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x), \tag{1.1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$. The problem (1.1.1) has various applications in the real world including science, engineering, informatics, economics and many others. Moreover, since it is the fundamental problem in nonlinear optimization problems, solution methods and theories for (1.1.1) provide the foundation to solve other nonlinear optimization problems with some constraints. Thus, it is very important to study on the unconstrained minimization problem. For solving (1.1.1), many solution methods, such as the steepest descent method and the Newton-type methods, have been developed. Recent advances in computer technologies enable us to solve some large-scale problems by these solution methods. However, there still exist a lot of problems that cannot be solved in practical time by the current state-of-the-art technologies.

When f is differentiable, the first-order necessary optimality condition for (1.1.1) is given by

$$\nabla f(x) = 0. \tag{1.1.2}$$

A point satisfying the condition is called a stationary point. When f is differentiable and convex, the point x^* is a stationary point if and only if x^* is an optimal solution. The main goal of solution methods for (1.1.1) is to find a stationary point.

Many problems are related to the unconstrained minimization problem (1.1.1). In what follows, we introduce the system of equations, the least squares problem, and the nonlinear complementarity problem (NCP) as the related problems, which are dealt with in this thesis. The system of equations is the problem to find $x \in \mathbb{R}^n$ such that

$$F(x) = 0, \tag{1.1.3}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$. The system (1.1.3) itself also has many applications in various fields. For example, the system (1.1.2) for a stationary point of f is regarded as (1.1.2) with $F = \nabla f$. Thus, the system (1.1.3) is strongly concerned with optimization problems. Moreover, as described afterward, the NCP can be reformulated to the system of equations.

The least squares problem is described as follows.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x) := \frac{1}{2}\|F(x)\|^2, \tag{1.1.4}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$, and $\|F(x)\|$ is the Euclidean norm of $F(x)$ defined by $\|F(x)\| := \sqrt{F(x)^T F(x)}$. The problem (1.1.4) often appears in data fitting problems [47, 54]. For $j = 1, \ldots, m$, let $\xi^j \in \mathbb{R}^l$ be input data, and $\eta_j \in \mathbb{R}$ be output data observed by using ξ^j. We estimate the relationship between the input and the output by a given fitting function $\phi(x, \xi)$ where $x \in \mathbb{R}^n$ is a parameter of ϕ. By defining $F(x)$ as

$$F(x) := \begin{pmatrix} \phi(x, \xi^1) - \eta_1 \\ \vdots \\ \phi(x, \xi^m) - \eta_m \end{pmatrix},$$

a solution x^* of the system (1.1.3) can be regarded as an optimal fitting parameter. Note that, when the system (1.1.3) has a solution, it is equivalent to the problem (1.1.4). Moreover, when $\nabla F(x^*)$ is row full rank, the point x^* is a stationary point of f if and only if x^* is a solution of (1.1.3). The objective function f of the problem (1.1.4) is called the least squares merit function. Thus, the problem (1.1.4) has the deep relationship with the system (1.1.3). Meanwhile, when the system (1.1.3) is overdetermined, i.e., $n < m$, it may have no solution. In such case, we assume that $\epsilon_j := \phi(x, \xi^j) - \eta_j$ has some estimate error or observation error. If ϵ_j's are independent and identically distributed, and follow a normal distribution with mean 0 and variance σ^2, then the maximum likelihood estimate of x is given as a solution of the following problem.

$$\underset{x \in \mathbb{R}^n}{\text{maximize}}\ p(x) := \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\phi(x, \xi^j) - \eta_j)^2}{2\sigma^2}\right) \tag{1.1.5}$$

The problem (1.1.5) is equivalent to the problem (1.1.4).

The NCP [13, 14, 18, 23, 24, 26, 32, 35, 40, 44], on which studies began in 1960s, is the problem to find $x \in \mathbb{R}^n$ such that

$$x \geq 0,\ G(x) \geq 0,\ x^T G(x) = 0,$$

where $G : \mathbb{R}^n \to \mathbb{R}^n$. The NCP includes important applications, such as traffic equilibrium problems [25, 27]. Moreover, it is closely related to constrained minimization problems. Consider the following problem.

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x), \\ &\text{subject to } x \geq 0. \end{aligned} \tag{1.1.6}$$

The Karush-Kuhn-Tucker (KKT) conditions for (1.1.6) are given by

$$\nabla f(x) - y = 0,\ x \geq 0,\ y \geq 0,\ x^T y = 0,$$

where y is the Lagrange multiplier. Thus, the problem (1.1.6) is equivalent to the NCP with $G = \nabla f$. The NCP can be reformulated to the system of equations as follows.

$$F(x) := \begin{pmatrix} \psi(x_1, G_1(x)) \\ \vdots \\ \psi(x_n, G_n(x)) \end{pmatrix} = 0,$$

where x_i is the i-th component of x, $G_i(x)$ is the i-th component of $G(x)$, and $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a NCP function such that

$$\psi(a, b) = 0 \iff a \geq 0,\ b \geq 0,\ ab = 0.$$

In 1992, Fischer [26] proposed the NCP function ψ defined by

$$\psi(a, b) := \sqrt{a^2 + b^2} - a - b,$$

which is called the Fischer-Burmeister function. Since ψ is not differentiable at $(0, 0)$, F is not differentiable at x if there exists i such that $x_i = G_i(x) = 0$. Thus, when we use the Fischer-Burmeister function, the NCP is reformulated to the system of nonsmooth equations.

1.2 Classical Newton method and Newton-type methods

The classical Newton method, which is first proposed by Newton in the 17th century, is one of the fundamental solution methods for solving the unconstrained minimization problem (1.1.1) or the system of equations (1.1.3). In what follows, we merely call the classical Newton method the Newton method.

First, we consider the problem (1.1.1), where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. Let x^k be the k-th iterative point. In the Newton method for (1.1.1), a search direction d^k is given as a solution of the system of linear equations

$$\nabla^2 f(x^k) d = -\nabla f(x^k), \tag{1.2.1}$$

and the next iterative point x^{k+1} is generated by $x^{k+1} := x^k + d^k$. In 1949, Kantorovich [34] analyzed the local convergence of the Newton method. If the second-order sufficient optimality condition holds at a solution x^*, i.e., $\nabla^2 f(x^*)$ is positive definite at x^*, and an initial point x^0 is sufficiently close to x^*, then the sequence $\{x^k\}$ converges to x^* quadratically, i.e., there exists a positive constant c such that $\|x^{k+1} - x^*\| \leq c\|x^k - x^*\|^2$.

Although the Newton method for (1.1.1) has rapid local convergence, the systems of linear equations (1.2.1) may have no solution when $\nabla^2 f(x^k)$ is singular. Moreover, when $\nabla^2 f(x^k)$ is not positive definite, the generated sequence $\{x^k\}$ does not necessarily converge to a stationary point. To resolve this drawback, some Newton-type methods, such as the quasi-Newton method, utilize the positive definite approximation of $\nabla^2 f(x^k)$. In these methods, a search direction d^k is well-defined and it is a descent direction for f at x^k, i.e., $\nabla f(x^k)^T d^k < 0$. Thus, by using an appropriate line search method, we can guarantee global convergence to a stationary point of f, i.e., $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$. In 1966, Armijo [1] proposed the line search strategy which is the most popular rule in this field. In Armijo's step size rule, we find the smallest nonnegative integer l_k such that

$$f(x^k) - f(x^k + \beta^{l_k} d^k) \geq -\alpha\beta^{l_k}\nabla f(x^k)^T d^k,$$

where α and β are positive constants such that $0 < \alpha < 1$ and $0 < \beta < 1$. Then, we set a step size t_k as $t_k := \beta^{l_k}$ and the next iterative point x^{k+1} as $x^{k+1} := x^k + t_k d^k$. Another technique to guarantee the global convergence of the Newton method is a trust-region method (trust-region NM) [17, 66], which have been eagerly studied since 1980s. Let Δ_k be a trust-region radius at the k-th iteration. In the trust-region NM, a search direction $d^k(\Delta_k)$ is given by a solution of a subproblem

$$\begin{aligned} &\underset{d\in\mathbb{R}^n}{\text{minimize}}\ \ f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d, \\ &\text{subject to } \|d\| \leq \Delta_k, \end{aligned} \tag{1.2.2}$$

and the next iterative point x^{k+1} is generated by $x^{k+1} := x^k + d^k(\Delta_k)$. For global convergence, the trust-region NM adaptively controls Δ_k instead of using a step size at each iteration. It is well-known that d is a solution of the subproblem (1.2.2) if and only if the following conditions hold [17].

$$\begin{aligned} &(\nabla^2 f(x^k) + \mu I)d = -\nabla f(x^k), \\ &\|d\| \leq \Delta_k, \quad \mu \geq 0, \quad (\|d\| - \Delta_k)\mu = 0, \\ &\mu \geq \max(0, -\lambda_{\min}(\nabla^2 f(x^k))), \end{aligned}$$

where $\mu \in \mathbb{R}$ is the Lagrange multiplier associated with the inequality constraint of the subproblem (1.2.2), and $\lambda_{\min}(\nabla^2 f(x^k))$ is the minimum eigenvalue of $\nabla^2 f(x^k)$. Thus, in the case where $\|d^k(\Delta_k)\| < \Delta_k$, $\mu = 0$ holds and hence the search direction $d^k(\Delta_k)$ is given as a solution of the system of linear equations (1.2.1). On the other hand, in the case where $\|d^k(\Delta_k)\| = \Delta_k$, the search direction $d^k(\Delta_k)$ is given by

$$d^k(\Delta_k) = -(\nabla^2 f(x^k) + \mu_k I)^{-1}\nabla f(x^k),$$

where μ_k is given as a solution of the nonlinear equation

$$\|(\nabla^2 f(x^k) + \mu I)^{-1}\nabla f(x^k)\| = \Delta_k, \quad \mu \geq \max(0, -\lambda_{\min}(\nabla^2 f(x^k))). \tag{1.2.3}$$

Since we have to solve the nonlinear equation (1.2.3), a lot of computational time may be required to get the search direction $d^k(\Delta_k)$. In Subsection 1.3.1, we will introduce the regularized Newton method whose subproblem is easier to solve as compared to (1.2.2).

Next, we consider the system (1.1.3), where $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable. In the Newton method for (1.1.3) with $n = m$, a search direction d^k is given as a solution of the system of linear equations

$$\nabla F(x^k) d = -F(x^k), \tag{1.2.4}$$

and the next iterative point x^{k+1} is generated by $x^{k+1} := x^k + d^k$. When $F = \nabla f$, the system (1.2.4) is equivalent to the system (1.2.1). Thus, the Newton method for (1.1.3) is equivalent to that for (1.1.1). If $\nabla F(x^*)$ is nonsingular at a solution x^* and an initial point x^0 is sufficiently close to x^*, then the sequence $\{x^k\}$ converges to x^* quadratically.

However, when $n \neq m$ or the Newton equation (1.2.4) may have no solution, we cannot apply the Newton method. In such cases, we may adopt the Gauss-Newton method. A search direction d^k of the Gauss-Newton method is given as a solution of the system of linear equations

$$\nabla F(x^k)^T \nabla F(x^k) d = -\nabla F(x^k)^T F(x^k). \tag{1.2.5}$$

The search direction d^k is also a solution of the least squares problem

$$\underset{d \in \mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|F(x^k) + \nabla F(x^k) d\|^2. \tag{1.2.6}$$

Thus, even if the system (1.1.3) has no solution, the least squares problem (1.2.6) always has a solution. However, when $\nabla F(x^k)$ is singular, d^k may not be a descent direction for the least squares merit function f at x^k. Therefore, in order to ensure convergence to a solution, the Gauss-Newton method requires an initial point to be sufficiently close to a solution. In Subsection 1.3.3, we will introduce the Levenberg-Marquardt method, which can overcome the drawback of the Gauss-Newton method.

1.3 Regularized Newton-type methods

1.3.1 Regularized Newton method

When f in the unconstrained minimization problem (1.1.1) is convex and twice continuously differentiable, the regularized Newton method (RNM) is one of the solution methods for (1.1.1) [38, 39, 57, 62, 63]. In the RNM, a search direction $d^k(\mu_k)$ is given by

$$d^k(\mu_k) := -(\nabla^2 f(x^k) + \mu_k I)^{-1} \nabla f(x^k),$$

where μ_k is a positive parameter. Since f is convex, $\nabla^2 f(x^k)$ is positive semidefinite, and it implies that $\nabla^2 f(x^k) + \mu_k I$ is positive definite. Thus, $d^k(\mu_k)$ is well-defined, and it is a descent direction for f at x^k. Therefore, the RNM with an appropriate line search method, such as Armijo's step size rule [1], has a global convergence property.

In the literatures [38, 39, 57], μ_k is set as $\mu_k := c\|\nabla f(x^k)\|$ for rapid local convergence, where c is a given positive constant. In 2004, Li, Fukushima, Qi and Yamashita [38] showed that the RNM with $\mu_k = c\|\nabla f(x^k)\|$ has a quadratic rate of convergence under a local error bound condition [56] (see Subsection 2.2.3 for the definition of the local error bound condition). Note that, the local error bound condition holds if the second-order sufficient optimality condition holds at x^*. But the converse is not true. Thus, the local error bound condition is weaker than the second-order sufficient optimality condition.

1.3.2 Cubic regularization of the Newton method

In 2006, Nesterov and Polyak [53] proposed the cubic regularization of the Newton method (cubic RNM) for the the unconstrained minimization problem (1.1.1) whose objective function f is twice continuously differentiable. The cubic RNM adopts a search direction $d^k(\sigma_k)$ as a solution of a subproblem

$$\underset{d\in\mathbb{R}^n}{\text{minimize}}\ f(x^k)+\nabla f(x^k)^T d+\frac{1}{2}d^T\nabla^2 f(x^k)d+\frac{1}{6}\sigma_k\|d\|^3, \tag{1.3.1}$$

where σ_k is a positive parameter. Note that, the cubic RNM can be applied even if f is nonconvex. Note also that, since a search direction of the steepest descent method is a solution of a subproblem

$$\underset{d\in\mathbb{R}^n}{\text{minimize}}\ f(x^k)+\nabla f(x^k)^T d+\frac{1}{2}\|d\|^2,$$

the cubic RNM is considered as a natural extension of the steepest descent method on the degree of approximation of f. Nesterov and Polyak [53] showed that d is a solution of the subproblem (1.3.1) if and only if the following conditions hold.

$$\left(\nabla^2 f(x^k)+\frac{1}{2}\sigma_k\|d\|I\right)d=-\nabla f(x^k),$$
$$\frac{1}{2}\sigma_k\|d\|\geq\max(0,-\lambda_{\min}(\nabla^2 f(x^k))).$$

Then, the search direction $d^k(\sigma_k)$ is given by

$$d^k(\sigma_k)=-\left(\nabla^2 f(x^k)+\frac{1}{2}\sigma_k r_k I\right)^{-1}\nabla f(x^k),$$

where r_k is given as a solution of the nonlinear equation

$$\left\|\left(\nabla^2 f(x^k)+\frac{1}{2}\sigma_k r I\right)^{-1}\nabla f(x^k)\right\|=r,\quad \frac{1}{2}\sigma_k r\geq\max(0,-\lambda_{\min}(\nabla^2 f(x^k))). \tag{1.3.2}$$

Since the nonlinear equation (1.3.2) is similar to (1.2.3), the subproblem (1.3.1) is difficult to solve as the subproblem (1.2.2) of the trust-region NM. Nesterov and Polyak also showed that the cubic RNM with the fixed parameter σ_k dependent of the Lipschitz constant of $\nabla^2 f$ has global convergence and quadratic rate of convergence under appropriate conditions.

Recently, Cartis, Gould and Toint [6, 8, 10, 12] extended the cubic RNM, called the adaptive cubic regularization method. For global convergence, the adaptive cubic regularization method adaptively controls the regularization parameter σ_k as the trust-region radius Δ_k of the trust-region NM.

1.3.3 Levenberg-Marquardt method

When F in the system (1.1.3) or the nonlinear least squares problem (1.1.4) is continuously differentiable, the Levenberg-Marqardt method (LMM) is one of the solution methods for (1.1.3) or (1.1.4) [5, 22, 33, 36, 37, 41, 42, 43, 45, 46, 54, 55, 59, 65, 64, 67, 68, 69]. The LMM is first proposed by Levenberg in 1944, and it is sometimes referred as the modified Gauss-Newton method. In the LMM, a search direction $d^k(\mu_k)$ is given by

$$d^k(\mu_k) := -(\nabla F(x^k)^T \nabla F(x^k) + \mu_k I)^{-1} \nabla F(x^k)^T F(x^k), \tag{1.3.3}$$

where μ_k is a positive parameter. We see that (1.3.3) is a regularized equation of the Gauss-Newton equation (1.2.5). The search direction d^k is also a solution of the following least squares problem.

$$\underset{d \in \mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|F(x^k) + \nabla F(x^k)d\|^2 + \frac{1}{2}\mu_k \|d\|^2 \tag{1.3.4}$$

Since $\nabla F(x^k)^T \nabla F(x^k) + \mu_k I$ is positive definite, $d^k(\mu_k)$ is well-defined even if $\nabla F(x^k)$ is singular. Moreover, from the positive definiteness of $\nabla F(x^k)^T \nabla F(x^k) + \mu_k I$, $d^k(\mu_k)$ is a descent direction for the least squares merit function f at x^k. In fact, if $\nabla f(x^k) \neq 0$, we have

$$\nabla f(x^k)^T d^k(\mu_k) = -F(x^k)^T \nabla F(x^k)(\nabla F(x^k)^T \nabla F(x^k) + \mu_k I)^{-1} \nabla F(x^k)^T F(x^k) < 0.$$

Thus, the LMM with an appropriate line search method has global convergence to a stationary point of f, and hence it is more useful as compared to the Gauss-Newton method.

Instead of using line search methods, we can also guarantee global convergence by updating μ_k appropriately [45, 55, 67]. In 1976, Osborne [55] proposed a simple and direct updating rule of μ_k, and showed the global convergence of the LMM with the rule. In 1978, Moré [45] proposed the updating rule based on the idea of the trust-region method [17, 66]. A search direction $d^k(\mu_k)$ of his proposal is given as a solution of a subproblem

$$\begin{aligned} &\underset{d \in \mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|F(x^k) + \nabla F(x^k)d\|^2, \\ &\text{subject to } \|d\|^2 \leq \Delta_k^2, \end{aligned} \tag{1.3.5}$$

and μ_k corresponds to the Lagrange multiplier of the KKT conditions of the subproblem. Then, instead of directly updating μ_k, Δ_k is controlled for global convergence. Thus, in his updating rule, we have to solve the subproblem (1.3.5) at each iteration. On the other hand, in Osborne's updating rule, we can get the search direction $d^k(\mu_k)$ by solving only the linear equations, which is much easier to solve than (1.3.5).

In addition to the global convergence property, the LMM has also rapid local convergence. In 2001, Yamashita and Fukushima [68] showed that the LMM with $\mu_k = \|F(x^k)\|^2$ has a quadratic rate of convergence under a local error bound condition [56].

When F is nonsmooth, we cannot use the Jacobian of F. Nevertheless, we can use the generalized Jacobian of F instead of ∇F if F is locally Lipschitz continuous [16, 58] (see Subsection 2.2.4 for the definitions of the generalized Jacobian and the locally Lipschitz continuity). In this thesis, we call the LMM using the generalized Jacobian of F the generalized LMM. The generalized LMM also has global and rapid local convergence properties under appropriate assumptions [22, 33].

1.4 Global complexity bound

Most of solution methods for the unconstrained minimization problem (1.1.1) are iterative methods which generate a sequence converging to a solution. Efficiencies of iterative methods are often discussed in terms of the global convergence or the rapid rate of convergence. The global convergence property ensures to get a solution from an arbitrary initial point. On the other hand, the rate of convergence represents the speed to get a solution from a point near the solution. However, the concept of global convergence is not related to the computational time, and the concept of rate of convergence ignores the time to get a point near a solution from an initial point. Thus, these concepts cannot provide a total computational time to get an appropriate solution from an arbitrary initial point. Recently, global complexity bounds of iterative methods for solving (1.1.1) have been actively investigated [6, 10, 29, 49, 50, 52, 53, 57]. The global complexity bound is defined as follows.

Definition 1.4.1. *Let ε_f be given positive constants. The global complexity bound of an iterative method with respect to the unconstrained minimization problem* (1.1.1) *is an upper bound of the number of iterations required to an approximate solution x satisfying*

$$f(x) - \inf_{y\in\mathbb{R}^n} f(y) \le \varepsilon_f. \tag{1.4.1}$$

Definition 1.4.2. *Let ε_g be given positive constants. The global complexity bound of an iterative method with respect to the stationary point* (1.1.2) *is an upper bound of the number of iterations required to an approximate solution x satisfying*

$$\|\nabla f(x)\| \le \varepsilon_g. \tag{1.4.2}$$

In Definitions 1.4.1 and 1.4.2, subscripts of ε_f and ε_g represent initial words of “function” and “gradient”, respectively. When f is nonconvex, the condition (1.4.2) is often used since it is difficult to estimate $\inf_{y\in\mathbb{R}^n} f(y)$. In what follows, when the global complexity bound is given with ε_f only, we discuss the bound with respect to (1.1.1), i.e., the condition (1.4.1). Otherwise, when the global complexity bound is given with ε_g only, we discuss the bound with respect to (1.1.2), i.e., the condition (1.4.2). Since the global complexity bound corresponds to the worst computational time, it is useful when we want to estimate in advance the time for solving a large-scale problem.

The steepest descent method adopts a search direction d^k as

$$d^k = -\nabla f(x^k).$$

Under the assumption that ∇f is Lipschitz continuous, Nesterov [50] showed that the global complexity bound of the steepest descent method is $O(\varepsilon_g^{-2})$ when f is nonconvex, and $O(\varepsilon_f^{-1})$ when f is convex. Moreover, the steepest descent method can be accelerated [49, 50]. The global complexity bound of the accelerated steepest descent method can be reduced to $O(\varepsilon_f^{-\frac{1}{2}})$.

For the trust-region NM, Gratton, Sartenaer and Toint [29] gave the global complexity bound.

They showed the bound is $O(\varepsilon_g^{-2})$ under the assumption that $\nabla^2 f$ is Lipschitz continuous. In their analysis, the complexity for solving the subproblems to get a search direction has not been considered.

For the RNM, the global complexity bound is also investigated. In 2009, Polyak [57] considered an RNM using a special step size with the Lipschitz constant of ∇f. The global complexity bound of his RNM is $O(\varepsilon_g^{-4})$ under the assumptions that f is convex and ∇f is Lipschitz continuous.

Recently, Nesterov and Polyak [53] showed that under the assumption that $\nabla^2 f$ is Lipschitz continuous, the global complexity bound of the cubic RNM is $O(\varepsilon_g^{-\frac{3}{2}})$ when f is nonconvex, and $O(\varepsilon_f^{-\frac{1}{2}})$ when f is convex. More recently, Cartis, Gould and Toint [6, 8, 10, 12] showed that the adaptive cubic RNM has the same complexities as the cubic RNM. For the unconstrained convex minimization problem, Nesterov [52] proposed an accelerated cubic RNM in a way similar to the acceleration technique for the steepest descent method and analyzed its global complexity bound. The global complexity bound of the accelerated cubic RNM can be reduced to $O(\varepsilon_f^{-\frac{1}{3}})$.

The existing results for the global complexity bound are summarized as the following table.

Table 1.1: The existing results for the global complexity bound

Method	Bound (nonconvex)	Bound (convex)	Lipschitz
Steepest descent method	$O(\varepsilon_g^{-2})$	$O(\varepsilon_f^{-1})$	∇f
Accelerated steepest descent method	–	$O(\varepsilon_f^{-\frac{1}{2}})$	∇f
Trust-region NM	$O(\varepsilon_g^{-2})$	–	$\nabla^2 f$
(Polyak's) RNM	–	$O(\varepsilon_g^{-4})$	∇f
Cubic RNM	$O(\varepsilon_g^{-\frac{3}{2}})$	$O(\varepsilon_f^{-\frac{1}{2}})$	$\nabla^2 f$
Accelerated cubic RNM	–	$O(\varepsilon_f^{-\frac{1}{3}})$	$\nabla^2 f$

1.5 Motivations and contributions

Various researches on solution methods for unconstrained minimization problems and their global complexity bounds have been done. However, there remain still many issues to be studied.

The trust-region NM and the cubic RNM have nice theoretical convergence properties. However, in order to get a search direction, these methods have to solve nonconvex subproblems at each iteration. Any global complexity bounds for solving these subproblems have not been analyzed. Thus, it is difficult to estimate the time for solving these subproblems. On the other hand, when f is convex, the RNM with line search can get a search direction by only solving linear equations. However, in most past studies for the RNM, the convergence properties have been discussed only when f is convex. Thus, we will propose an RNM extended to the problem (1.1.1), whose objective function f is nonconvex. We will show that the extended RNM with line search has good convergence properties, though f is nonconvex. Moreover, we will investigate its global complexity bound.

From various numerical experiments, it is known that the number of function evaluations of the trust-region NM required to get a solution tends to be less as compared to that of the Newton-type methods with some line search method. Therefore, it is desirable to construct a solution method whose behavior is similar to the trust-region NM, and subproblems can be solved as the RNM. Thus, we will propose an RNM without line search, which controls the regularization parameter for global convergence. We will show its good convergence properties and give its global complexity bound.

For solving the nonlinear least squares problem (1.1.4), we can also apply the steepest descent method and the Newton-type methods. From the results of the global complexity bounds, if we apply the steepest descent method or the Newton-type methods to (1.1.4), then we can estimate the worst computational time in advance. However, since these methods are not specialized to (1.1.4), they are not efficient. In fact, the steepest descent method converges slow in general, and the Newton-type methods require the twice continuous differentiability of F. Thus, it is worth investigating the global complexity bound for methods specified for (1.1.4). The LMM and the generalized LMM are the special methods for (1.1.4), and the global complexity bounds for the LMM and the generalized LMM remains unknown. Therefore, we will investigate the global complexity bounds for these methods in this thesis.

1.6 Outline of the thesis

This thesis is organized as follows.

In Chapter 2, we give some mathematical notations and concepts which are used in this thesis.

In Chapter 3, we propose an RNM extended to the unconstrained nonconvex minimization problem (1.1.1). We show that the extended RNM has global convergence and rapid rate of convergence under appropriate conditions. We also show that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumption that $\nabla^2 f$ is Lipschitz continuous.

In Chapter 4, we propose an RNM without line search. To guarantee global convergence, the proposed method controls μ_k directly in the same way as Osborne's updating rule [55] for the LMM. We show that the proposed method has global convergence and rapid rate of convergence under appropriate conditions. Moreover, under the assumption that $\nabla^2 f$ is Lipschitz continuous, we show that the the global complexity bounds are $O(\varepsilon_g^{-2})$ when f is nonconvex, $O(\varepsilon_g^{-\frac{5}{3}})$ and $O(\varepsilon_f^{-\frac{2}{3}})$ when f is convex, and $O(\varepsilon_g^{-1})$ and $O(\log \varepsilon_f^{-1})$ when f is strongly convex.

In Chapter 5, we investigate the global complexity bound of the LMM with Osborne's updating rule [55] for solving the nonlinear least squares problem (1.1.4). We show that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumption that ∇F is Lipschitz continuous.

In Chapter 6, we investigate the global complexity bound of the generalized LMM for the system of nonsmooth equations (1.1.3). By reformulating the system (1.1.3) to the nonlinear least squares problem (1.1.4), we show that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumption that ∇f is Lipschitz continuous. We also show that the global complexity bound is $O(\log \varepsilon_f^{-1})$ under some regularity assumption. Furthermore, by applying these results to nonsmooth equations equivalent to the NCP, we analyze the global complexity bounds for the NCP.

Finally, Chapter 7 summarizes the results obtained in this thesis, and mentions some possible future works.

Chapter 2

Preliminaries

In this chapter, we give mathematical notations and concepts that will be used in the subsequent chapters.

2.1 Notations

Throughout this thesis, we use the following notations. For a vector $x \in \mathbb{R}^n$, x_i denotes the i-th component of x. Then, $\|x\|$ denotes the Euclidean norm defined by

$$\|x\| := \sqrt{x^T x} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

For a vector mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, $F_i(x)$ denotes the i-th component of $F(x)$. For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, we denote the maximum eigenvalue and the minimum eigenvalue of M as $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$, respectively. Then, for a matrix $M \in \mathbb{R}^{n\times m}$, $\|M\|$ denotes the ℓ_2 norm of M defined by

$$\|M\| := \max_{v\in\mathbb{R}^m,\ \|v\|=1} \|Mv\| = \sqrt{\lambda_{\max}(M^T M)}.$$

Note that, if M is a symmetric positive semidefinite matrix, then

$$\|M\| = \lambda_{\max}(M).$$

Furthermore, for a symmetric matrix $M \in \mathbb{R}^{n\times n}$, $M \succ (\succeq) 0$ denotes the positive (semi)definiteness of M, i.e., $\lambda_{\min}(M) > (\geq) 0$. $B(x, r)$ denotes the closed sphere with center $x \in \mathbb{R}^n$ and radius $r \geq 0$, i.e.,

$$B(x, r) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}.$$

$\mathrm{dist}(x, S)$ denotes the distance between a vector $x \in \mathbb{R}^n$ and a set $S \subseteq \mathbb{R}^n$, i.e.,

$$\mathrm{dist}(x, S) := \min_{y\in S} \|y - x\|.$$

Moreover, we denote co$S$ and $\mathcal{P}(S)$ as the convex hull of S and the set consisting of all the subsets of S, respectively. For sets $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^n$, $S_1 + S_2$ denotes the sum of S_1 and S_2 defined by

$$S_1 + S_2 := \{x + y \in \mathbb{R}^n \mid x \in S_1, y \in S_2\}.$$

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, its gradient $\nabla f(x)$ is defined by

$$\nabla f(x) := \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix},$$

where $\partial f(x)/\partial x_i$ denotes the partial derivative of f at x associated with its i-th component. In addition, when f is twice differentiable, its Hessian matrix $\nabla^2 f(x)$ is defined by

$$\nabla^2 f(x) := \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

In what follows, x^k denotes the k-th iterative point generated by an iterative method. We also denote $f(x^k)$, $\nabla f(x^k)$ and $\nabla^2 f(x^k)$ as f_k, g^k and H_k, respectively. For a differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, its Jacobian matrix $\nabla F(x)$ is defined by

$$\nabla F(x) := \begin{pmatrix} \nabla F_1(x)^T \\ \vdots \\ \nabla F_m(x)^T \end{pmatrix} = \begin{pmatrix} \dfrac{\partial F_1(x)}{\partial x_1} & \cdots & \dfrac{\partial F_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m(x)}{\partial x_1} & \cdots & \dfrac{\partial F_m(x)}{\partial x_n} \end{pmatrix}.$$

In what follows, we denote $F(x^k)$ and $\nabla F(x^k)$ as F^k and J_k, respectively.

2.2 Definitions

2.2.1 Lipschitz continuity

Lipschitz continuity plays an important role in achieving rapid local convergence and estimating global complexity bounds. We first give the definitions of Lipschitz continuity.

Definition 2.2.1. *Let S be a subset of $\mathbb{R}^n$, $f : S \to \mathbb{R}$ be a twice differentiable function, and $F : S \to \mathbb{R}^m$ be a differentiable mapping.*

(a) *f is said to be Lipschitz continuous (on S) if there exists a positive constant L_f such that*

$$|f(x) - f(y)| \le L_f \|x - y\|, \quad \forall x, y \in S.$$

(b) ∇f *is said to be Lipschitz continuous* (*on* S) *if there exists a positive constant* L_g *such that*

$$\|\nabla f(x) - \nabla f(y)\| \leq L_g \|x - y\|, \quad \forall x, y \in S.$$

(c) $\nabla^2 f$ *is said to be Lipschitz continuous* (*on* S) *if there exists a positive constant* L_H *such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_H \|x - y\|, \quad \forall x, y \in S.$$

(d) F *is said to be Lipschitz continuous* (*on* S) *if there exists a positive constant* L_F *such that*

$$\|F(x) - F(y)\| \leq L_F \|x - y\|, \quad \forall x, y \in S.$$

(e) ∇F *is said to be Lipschitz continuous* (*on* S) *if there exists a positive constant* L_J *such that*

$$\|\nabla F(x) - \nabla F(y)\| \leq L_J \|x - y\|, \quad \forall x, y \in S.$$

The constants L_f, L_g, L_H, L_F and L_J are called Lipschitz constant.

When $f : \mathbb{R}^n \to \mathbb{R}$ is an affine function, i.e., f is defined by $f(x) := b^T x + c$, f is Lipschitz continuous with modulus $L_f = \|b\|$. Moreover, ∇f and $\nabla^2 f$ are also Lipschitz continuous with modulus $L_g = 0$ and $L_H = 0$. On the other hand, when $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function, i.e., f is defined by $f(x) := \frac{1}{2} x^T A x + b^T x + c$, f is not Lipschitz continuous on $\mathbb{R}^n$. However, ∇f and $\nabla^2 f$ are Lipschitz continuous with modulus $L_g = \|A\|$ and $L_H = 0$.

The following proposition shows the relations between Lipschitz continuity and compactness of a set. It implies that many classes of continuously differentiable functions are Lipschitz continuous on a compact set.

Proposition 2.2.1. *Let* S *be a compact set in* $\mathbb{R}^n$, $f : S \to \mathbb{R}$ *be a function, and* $F : S \to \mathbb{R}^m$ *be a mapping. Then,*

(a) *If* f *is continuously differentiable, then* f *is Lipschitz continuous on* S.

(b) *If* f *is twice continuously differentiable, then* ∇f *are Lipschitz continuous on* S.

(c) *If* F *is differentiable, then* F *is Lipschitz continuous on* S.

The following proposition shows the relations between Lipschitz continuity and uniformly boundedness of derivative.

Proposition 2.2.2. *Let* S *be a subset of* $\mathbb{R}^n$, $f : S \to \mathbb{R}$ *be a twice differentiable function,* $F : S \to \mathbb{R}^m$ *be a differentiable mapping. Then,*

(a) f *is Lipschitz continuous* (*on* S) *with modulus* L_f *if and only if*

$$\|\nabla f(x)\| \leq L_f, \quad \forall x \in S.$$

(b) ∇f *is Lipschitz continuous* (*on* S) *with modulus* L_g *if and only if*

$$\|\nabla^2 f(x)\| \leq L_g, \quad \forall x \in S.$$

(c) F *is Lipschitz continuous* (*on* S) *with modulus* L_F *if and only if*

$$\|\nabla F(x)\| \leq L_F, \quad \forall x \in S.$$

2.2.2 Convexity

We define the convexity of a function as follows.

Definition 2.2.2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be*

(a) *convex if*

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y), \quad \forall x, y \in \mathbb{R}^n, \quad \forall \alpha \in (0,1).$$

(b) *strongly convex if there exists a positive constant σ such that*

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) - \frac{1}{2}\sigma(1-\alpha)\alpha\|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n, \quad \forall \alpha \in (0,1).$$

It is obvious that any strongly convex function is convex. But a convex function is not necessarily strongly convex. For example, a function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := x^4$ is convex but not strongly convex.

The following proposition mentions the relations between the convexity of a function and the positive (semi)definiteness of its Hessian matrix.

Proposition 2.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Then,*

(a) *f is convex if and only if*

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbb{R}^n.$$

(b) *f is strongly convex if and only if there exists a positive constant σ such that*

$$\min_{v \in \mathbb{R}^n,\ \|v\|=1} v^T \nabla^2 f(x) v \geq \sigma, \quad \forall x \in \mathbb{R}^n.$$

2.2.3 Local error bound

The concept of a local error bound is important for rapid local convergence in a neighborhood of a solution [56]. A local error bound used in this thesis is defined as follows.

Definition 2.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and X^* be the local optimal solution set of the unconstrained minimization problem* (1.1.1)*. $\|\nabla f(x)\|$ provides a local error bound for* (1.1.1) *if for any solution x^* of* (1.1.1)*, there exist positive constants b and κ such that*

$$\kappa \operatorname{dist}(x, X^*) \leq \|\nabla f(x)\|, \quad \forall x \in B(x^*, b).$$

The following proposition shows that the local error bound condition is a generalization of the second-order sufficient optimality condition.

Proposition 2.2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function, and x^* be a local optimal solution of* (1.1.1)*. If the second-order sufficient optimality condition holds at x^*, i.e., there exists a positive constant σ such that $\lambda_{\min}(\nabla^2 f(x^*)) \geq \sigma$, then $\|\nabla f(x)\|$ provides a local error bound for* (1.1.1)*.*

Note that, the converse of the above proposition is not true. In fact, we consider the case where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x_1, x_2) = x_1^2$. Then, $\|\nabla f(x)\| = 2|x_1|$. The local optimal solution set X^* is given by $X^* = \{x \in \mathbb{R}^2 \mid x_1 = 0\}$. It then follows that $\mathrm{dist}(x, X^*) = |x_1|$. Thus, the local error bound condition holds for any $\kappa \in (0, 0.5]$ and $b \in (0, \infty)$. However, since $\nabla^2 f(x)$ is singular for any $x \in X^*$, the second-order sufficient optimality condition does not hold. Therefore, the local error bound condition is weaker than the second-order sufficient optimality condition.

2.2.4 Generalized Jacobian

We give some concepts which extend the idea of differentiability to nondifferentiable mapping. When a vector mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is nonsmooth, we cannot necessarily use the Jacobian of F. Nevertheless, if F is locally Lipschitz continuous, i.e., for any bounded set $\Omega \subset \mathbb{R}^n$, there exists a positive constant $L_F(\Omega)$ such that

$$\|F(x) - F(y)\| \le L_F(\Omega)\|x - y\|, \quad \forall x, y \in \Omega,$$

then, we can define the generalized Jacobian of F [16, 58].

Definition 2.2.4. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous mapping, and $D_F \subseteq \mathbb{R}^n$ be the set where F is differentiable. The point-to-set mappings $\partial_B F$ and $\partial F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^{m \times n})$ defined by*

$$\begin{aligned} \partial_B F(x) &:= \{J \in \mathbb{R}^{n \times m} \mid J = \lim_{k \to \infty} \nabla F(y^k),\ \lim_{k \to \infty} y^k = x,\ \{y^k\} \subseteq D_F\} \\ \partial F(x) &:= \mathrm{co}\, \partial_B F(x) \end{aligned}$$

are said to be the B-subdifferential and the Clarke generalized Jacobian of F at x, respectively.

In what follows, we merely call the Clarke generalized Jacobian the generalized Jacobian. When $F : \mathbb{R} \to \mathbb{R}$ is defined by $F(x) := |x|$, we have $\partial_B F(0) = \{-1, 1\}$ and hence $\partial F(0) = [-1, 1]$. The generalized Jacobian is a natural extension of the original Jacobian for differential functions. In fact, if F is continuously differentiable at x, then we have $\partial_B F(x) = \partial F(x) = \{\nabla F(x)\}$.

Note that, if F is locally Lipschitz continuous, then F is differentiable almost everywhere, and $\partial_B F(x)$ and $\partial F(x)$ are nonempty and compact sets for each x [16]. Moreover, if $f(x) := \frac{1}{2}\|F(x)\|^2$ is continuously differentiable, we have $\nabla f(x) = J^T F(x),\ \forall J \in \partial F(x)$ by using the standard calculus rules [16].

Since the generalized Jacobian is a point-to-set mapping, we introduce concepts related to continuity of a point-to-set mapping. The upper semi-continuity defined below is useful in analysis using the generalized Jacobian [30].

Definition 2.2.5. *Let X be a subset of $\mathbb{R}^n$, Y be a subset of $\mathbb{R}^{m \times n}$, and Θ be a point-to-set mapping from X into $\mathcal{P}(Y)$.*

(a) *Θ is said to be uniformly compact near $\bar{x} \in X$ if there exists a neighborhood N of $\bar{x}$ such that the closure of $\cup_{x \in N} \Theta(x)$ is compact.*

(b) *Θ is said to be closed at $\bar{x}$ if $x^k \to \bar{x}, y^k \in \Theta(x^k)$ and $y^k \to \bar{y}$ imply $\bar{y} \in \Theta(\bar{x})$.*

(c) *Θ is said to be upper semi-continuous at $\bar{x}$ if Θ is uniformly compact near $\bar{x}$ and closed at $\bar{x}$.*

For example, when a point-to-set mapping $\Theta(x) : \mathbb{R} \to \mathcal{P}(\mathbb{R})$ is defined by

$$\Theta(x) := \begin{cases} \{y \in \mathbb{R} \mid 0 \leq y \leq 1\} & (x \leq 0), \\ \{y \in \mathbb{R} \mid 0 \leq y \leq 1/x\} & (x > 0), \end{cases}$$

the mapping Θ is upper semi-continuous at 1. However, Θ is neither uniformly compact near 0 nor closed at 0, and hence Θ is not upper semi-continuous at 0. If Θ is a continuous mapping, then Θ is upper semi-continuous for each x.

It is well-known that for a locally Lipschitz continuous mapping F, ∂F is upper semi-continuous [16]. Thus, for each x, $\max_{J \in \partial F(x)} \|J\|$ is bounded from above.

Chapter 3

A regularized Newton method with line search for the unconstrained nonconvex minimization problem and its convergence properties

3.1 Introduction

In this chapter, we consider the following unconstrained minimization problem.

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f(x), \tag{3.1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function.

When f is convex, the regularized Newton method (RNM) is one of the efficient solution methods for (3.1.1) and has good convergence properties [38, 39, 57]. For the k-th iterative point x^k, the RNM adopts a search direction $d^k(\mu_k)$ defined by

$$d^k(\mu_k) := -(H_k + \mu_k I)^{-1} g^k,$$

where μ_k is a positive parameter. If f is convex, then $H_k + \mu_k I$ is a positive definite matrix, and hence $d^k(\mu_k)$ is a descent direction for f at x^k. Therefore, the RNM with an appropriate line search method, such as Armijo's step size rule, has a global convergence property.

In the literatures [38, 39, 57], μ_k is set as $\mu_k := c\|\nabla f(x^k)\|$ for rapid local convergence, where c is a given positive constant. Li, Fukushima, Qi and Yamashita [38] showed that the RNM has a quadratic rate of convergence under the assumption that $\|\nabla f(x)\|$ provides a local error bound for (3.1.1). Recently, Polyak [57] proposed the RNM with a special step size

$$t_k = \frac{\|g^k\|}{L_g}, \tag{3.1.2}$$

where L_g denotes the Lipschitz constant of ∇f. He showed that the global complexity bound of his method is

$$O(\varepsilon_g^{-4})$$

without assuming the local error bound condition nor the second-order sufficient optimality condition. However, Polyak's RNM uses a special step size which includes the Lipschitz constant. Thus, the above result may not hold if the Lipschitz constant is unknown.

In most past studies for the RNM, the convergence properties have been discussed only when f is convex. In this chapter, we consider an RNM extended to the problem (3.1.1) whose objective function f is nonconvex. The extended RNM uses Armijo's step size rule, and it does not contain unknown constants, e.g., the Lipschitz constant of ∇f as Polyak's method. We show that the extended RNM has the following properties:

- If the level set Ω of f at the initial point x^0 is compact, then $\|g^k\|$ converges to 0.
- If the level set Ω is compact and $\nabla^2 f$ is Lipschitz continuous on a certain compact set containing Ω, then the global complexity bound of the extended RNM is

$$O(\varepsilon_g^{-2}).$$

- Under the local error bound condition, the distance between x^k and the local optimal solution set X^* converges to 0 superlinearly, i.e., there exists a sequence $\{\alpha_k\}$ such that

$$\mathrm{dist}(x^{k+1}, X^*) \leq \alpha_k \mathrm{dist}(x^k, X^*), \quad \lim_{k\to\infty} \alpha_k = 0.$$

This chapter is organized as follows. In the next section, we extend the RNM to the problem (3.1.1) whose objective function f is not necessarily convex. In Section 3.3, we show that the extended RNM has global convergence. In Section 3.4, we give the global complexity bound of the extended RNM. In Section 3.5, we establish superlinear convergence under the local error bound condition. Finally, Section 3.6 concludes this chapter.

3.2 Proposed algorithm

In this section, we extend the RNM to the unconstrained nonconvex minimization problem (3.1.1). At the k-th iteration of the RNM, we set the regularization parameter μ_k as

$$\mu_k := c_1\Lambda_k + c_2\|g^k\|^\delta,$$

where c_1, c_2, δ are given constants such that $c_1 > 1, c_2 > 0, \delta \geq 0$, and Λ_k is defined by

$$\Lambda_k := \max(0, -\lambda_{\min}(H_k)).$$

From the definition of Λ_k, the matrix $H_k + \Lambda_k I$ is positive semidefinite even if f is nonconvex. Therefore, if $\|g^k\| \neq 0$, then $H_k + \mu_k I = H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I \succ 0$. Thus, we can adopt a search direction $d^k(\mu_k)$ at x^k as

$$d^k(\mu_k) = -(H_k + \mu_k I)^{-1} g^k = -(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} g^k$$

when f is nonconvex. In what follows, we merely denote the search direction as d^k instead of $d^k(\mu_k)$.

The algorithm of the RNM with the above d^k and Armijo's step size rule is described as follows. In what follows, we call the proposed algorithm the extended RNM.

The Extended Regularized Newton Method

Step 0 : Choose parameters $\delta, c_1, c_2, \alpha, \beta$ such that

$$\delta \geq 0,\ c_1 > 1,\ c_2 > 0,\ 0 < \alpha < 1,\ 0 < \beta < 1.$$

Choose a starting point x^0. Set $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : Compute d^k.

Step 3 : Find the smallest nonnegative integer l_k such that

$$f_k - f(x^k + \beta^{l_k} d^k) \geq -\alpha \beta^{l_k} {g^k}^T d^k. \tag{3.2.1}$$

Step 4 : Set $t_k := \beta^{l_k}$, $x^{k+1} := x^k + t_k d^k$ and $k := k + 1$. Go to Step 1.

In Step 3 of the extended RNM, a backtracking scheme is used. Since ${g^k}^T d^k < 0$ for k such that $\|g^k\| \neq 0$, the number of backtracking steps is finite, that is, the extended RNM is well-defined.

3.3 Global convergence

In this section, we show that the extended RNM has global convergence to a stationary point. To this end, we need the following assumption.

Assumption 3.3.1. *The level set of f at the initial point x^0 is compact, i.e., $\Omega := \{x \in \mathbb{R}^n \mid f(x) \leq f_0\}$ is compact.*

Since $\{f_k\}$ is monotonically decreasing, the sequence $\{x^k\}$ is included in the compact set Ω under Assumption 3.3.1. Then, there exists $f_{\min}$ such that

$$f_k \geq f_{\min}, \quad \forall k \geq 0. \tag{3.3.1}$$

Moreover, since f is twice continuously differentiable, f is Lipschitz continuous on Ω. Thus, there exists a positive constant L_f such that

$$\|g^k\| \leq L_f, \quad \forall k \geq 0. \tag{3.3.2}$$

First, we give an upper bound of $\|d^k\|$.

Lemma 3.3.1. *Suppose that* $\|g^k\| \neq 0$. *Then,*

$$\|d^k\| \leq \frac{\|g^k\|^{1-\delta}}{c_2}.$$

Proof. From the definition of d^k, we have

$$\begin{aligned}
\|d^k\| &= \|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} g^k\| \\
&\leq \|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\| \cdot \|g^k\| \\
&= \lambda_{\max}\Big((H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\Big)\|g^k\| \\
&= \frac{\|g^k\|}{\lambda_{\min}(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)} \\
&= \frac{\|g^k\|}{(c_1 - 1)\Lambda_k + c_2\|g^k\|^\delta} \\
&\leq \frac{\|g^k\|^{1-\delta}}{c_2},
\end{aligned}$$

which is the desired inequality. □

The next lemma indicates that $\|d^k\|$ is bounded from above if $\|g^k\|$ does not converge to 0.

Lemma 3.3.2. *Suppose that Assumption* 3.3.1 *holds. Suppose also that there exists a positive constant* ε_g *such that* $\|g^k\| \geq \varepsilon_g$. *Then,*

$$\|d^k\| \leq U_d(\varepsilon_g),$$

where

$$U_d(\varepsilon_g) := \max\left(\frac{L_f^{1-\delta}}{c_2}, \frac{1}{c_2\varepsilon_g^{\delta-1}}\right).$$

Proof. When $\delta \leq 1$, it follows from Lemma 3.3.1 and (3.3.2) that

$$\|d^k\| \leq \frac{L_f^{1-\delta}}{c_2}.$$

Meanwhile, when $\delta > 1$, it follows from Lemma 3.3.1 and $\|g^k\| \geq \varepsilon_g$ that

$$\|d^k\| \leq \frac{1}{c_2\varepsilon_g^{\delta-1}}.$$

This completes the proof. □

When there exists a positive constant ε_g such that $\|g^k\| \geq \varepsilon_g$ for all k, it follows from Lemma 3.3.2 that

$$x^k + \tau d^k \in \Omega + B(0, U_d(\varepsilon_g)), \quad \forall \tau \in [0,1], \quad \forall k \geq 0.$$

Moreover, since $\Omega + B(0, U_d(\varepsilon_g))$ is compact and f is twice continuously differentiable, ∇f is Lipschitz continuous on $\Omega + B(0, U_d(\varepsilon_g))$. Thus, there exists a positive constant $L_g(\varepsilon_g)$ such that

$$\|\nabla^2 f(x)\| \leq L_g(\varepsilon_g), \quad \forall x \in \Omega + B(0, U_d(\varepsilon_g)). \tag{3.3.3}$$

Now, we show that the condition (3.2.1) in Armijo's step size rule holds if $\|g^k\| \geq \varepsilon_g$ and the step size t_k is less than a specific value depending on ε_g.

Lemma 3.3.3. *Suppose that Assumption* 3.3.1 *holds, and there exists a positive constant* ε_g *such that* $\|g^k\| \geq \varepsilon_g$. *Suppose also that*

$$t_k \leq \frac{2(1-\alpha)c_2\varepsilon_g^\delta}{L_g(\varepsilon_g)}.$$

Then,

$$f_k - f(x^k + t_k d^k) \geq -\alpha t_k {g^k}^T d^k.$$

Proof. Since f is twice continuously differentiable, we have

$$f(x^k + t_k d^k) = f_k + t_k {g^k}^T d^k + \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau.$$

It then follows from the definition of d^k that

$$\begin{aligned}
&f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \\
&= -(1-\alpha) t_k {g^k}^T d^k - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau \\
&= (1-\alpha) t_k {d^k}^T (H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I) d^k - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau.
\end{aligned}$$

Since $H_k + c_1 \Lambda_k I$ is positive semidefinite, we have

$$\begin{aligned}
f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k &\geq \frac{1}{2} t_k {d^k}^T \left(2(1-\alpha) c_2 \|g^k\|^\delta I - t_k \int_0^1 \nabla^2 f(x^k + \tau t_k d^k) \mathrm{d}\tau \right) d^k \\
&\geq \frac{1}{2} t_k \|d^k\|^2 \left(2(1-\alpha) c_2 \|g^k\|^\delta - t_k \int_0^1 \|\nabla^2 f(x^k + \tau t_k d^k)\| \mathrm{d}\tau \right) \\
&\geq \frac{1}{2} t_k \|d^k\|^2 \left(2(1-\alpha) c_2 \varepsilon_g^\delta - t_k L_g(\varepsilon_g) \right) \\
&\geq 0,
\end{aligned}$$

where the third inequality follows from (3.3.3) and the assumption that $\|g^k\| \geq \varepsilon_g$, and the last inequality follows from the assumption on t_k. $\square$

From Lemma 3.3.3, we show that the step size t_k determined in Step 4 of the extended RNM is bounded away from 0 when $\|g^k\| \geq \varepsilon_g$.

Lemma 3.3.4. *Suppose that Assumption* 3.3.1 *holds. Suppose also that there exists a positive constant ε_g such that $\|g^k\| \geq \varepsilon_g$. Then,*

$$t_k \geq t_{\min}(\varepsilon_g),$$

where

$$t_{\min}(\varepsilon_g) := \min\left(1, \frac{2(1-\alpha)\beta c_2 \varepsilon_g^\delta}{L_g(\varepsilon_g)}\right).$$

Proof. From Lemma 3.3.3 and the definition of t_k in Step 4 of the extended RNM, t_k must be

$$t_k \geq \min\left(1, \frac{2(1-\alpha)\beta c_2 \varepsilon_g^\delta}{L_g(\varepsilon_g)}\right).$$

Otherwise βt_k satisfies Armijo's rule, which contradicts the definition of t_k. □

Next, we give a lower bound of the reduction $f_k - f_{k+1}$ when $\|g^k\| \geq \varepsilon_g$.

Lemma 3.3.5. *Suppose that Assumption* 3.3.1 *holds. Suppose also that there exists a positive constant ε_g such that $\|g^k\| \geq \varepsilon_g$. Then,*

$$f_k - f_{k+1} \geq p(\varepsilon_g)\varepsilon_g^2,$$

where

$$p(\varepsilon_g) := \frac{\alpha t_{\min}(\varepsilon_g)}{(1+c_1)L_g(\varepsilon_g) + c_2 L_f^\delta}.$$

Proof. From Armijo's rule and the definition of d^k, we have

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha t_k {g^k}^T d^k \\ &= \alpha t_k {g^k}^T (H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I)^{-1} g^k \\ &\geq \alpha t_k \lambda_{\min}\Big((H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I)^{-1}\Big)\|g^k\|^2. \end{aligned} \tag{3.3.4}$$

Since $H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I$ is positive definite, we have

$$\begin{aligned} \lambda_{\min}\Big((H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I)^{-1}\Big) &= \frac{1}{\lambda_{\max}(H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I)} \\ &= \frac{1}{\lambda_{\max}(H_k) + c_1 \Lambda_k + c_2 \|g^k\|^\delta} \\ &\geq \frac{1}{(1+c_1)L_g(\varepsilon_g) + c_2 L_f^\delta}, \end{aligned}$$

where the last inequality follows from (3.3.2) and (3.3.3). It then follows from (3.3.4) that

$$\begin{aligned} f_k - f_{k+1} &\geq \frac{\alpha t_k}{(1+c_1)L_g(\varepsilon_g) + c_2 L_f^\delta}\|g^k\|^2 \\ &\geq \frac{\alpha t_{\min}(\varepsilon_g)}{(1+c_1)L_g(\varepsilon_g) + c_2 L_f^\delta}\varepsilon_g^2, \end{aligned}$$

where the last inequality follows from Lemma 3.3.4 and the assumption on $\|g^k\|$. □

From the above lemma, we show the global convergence of the extended RNM.

Theorem 3.3.1. *Suppose that Assumption* 3.3.1 *holds. Then,*

$$\lim_{k\to\infty} \|g^k\| = 0.$$

Proof. Suppose the contrary, i.e., $\limsup_{k\to\infty} \|g^k\| > 0$. Let

$$\begin{aligned} &\varepsilon_g := \frac{1}{2}\limsup_{k\to\infty} \|g^k\|, \\ &K(k,\varepsilon_g) := \{j \in \{0,1,\dots\} \mid j \leq k,\ \|g^j\| \geq \varepsilon_g\}. \end{aligned}$$

Then, we have

$$\lim_{k\to\infty} |K(k,\varepsilon_g)| = \infty,$$

where $|K(k,\varepsilon_g)|$ denotes the number of the elements of $K(k,\varepsilon_g)$. From Lemma 3.3.5, we have

$$\begin{aligned} f_0 - f_{k+1} &\geq \sum_{j=0}^{k}(f_j - f_{j+1}) \\ &\geq \sum_{j\in K(k,\varepsilon_g)} (f_j - f_{j+1}) \\ &\geq \sum_{j\in K(k,\varepsilon_g)} p(\varepsilon_g)\varepsilon_g^2 \\ &= p(\varepsilon_g)\varepsilon_g^2 \left|K(k,\varepsilon_g)\right|. \end{aligned}$$

Taking $k \to \infty$, the right hand side of the inequality goes to infinity, and hence $\lim_{k\to\infty} f_k = -\infty$. This contradicts (3.3.1). Hence, we have $\limsup_{k\to\infty} \|g^k\| = 0$, i.e., $\lim_{k\to\infty} \|g^k\| = 0$. □

Remark 3.3.1. *Note that we can show the global convergence under a weaker condition that* $\{x^k\}$ *is included in some compact set.*

3.4 Global complexity bound

In this section, we estimate the global complexity bound of the extended RNM. Let $K_{\rm outer}$ be the total number of outer iterations until $\|g^k\| \le \varepsilon_g$ holds for the first time. If there does not exist such $K_{\rm outer}$, we define $K_{\rm outer} := \infty$. Moreover, let $K_{\rm total}$ be the total number of inner iterations until $k = K_{\rm outer}$ holds, i.e.,

$$K_{\rm total} := \sum_{k=0}^{K_{\rm outer}-1} (l_k + 1).$$

Note that, $K_{\rm outer}$ means the total number of solving linear equations, and $K_{\rm total}$ means the total number of function evaluation.

To investigate the global complexity bound, we need the following assumptions in addition to Assumption 3.3.1.

Assumption 3.4.1.

(a) $\delta \le 1/2$.

(b) $\alpha \le 1/2$.

(c) *Let* $b_1 := L_f^{1-\delta}/c_2$. $\nabla^2 f$ *is Lipschitz continuous on* $\Omega + B(0, b_1)$ *with modulus* L_H.

First, we give an upper bound of $\|d^k\|$.

Lemma 3.4.1. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then,*

$$\|d^k\| \le \frac{L_f^{1-\delta}}{c_2}.$$

Proof. Since $\delta \le 1$, it directly follows from Lemma 3.3.1 and (3.3.2). □

From Lemma 3.4.1, we have

$$x^k + \tau d^k \in \Omega + B(0, b_1), \quad \forall \tau \in [0, 1], \quad \forall k \ge 0.$$

It then follows from Assumption 3.4.1 (c) and $t_k \le 1$ that

$$\|\nabla^2 f(x^k + \tau t_k d^k) - H_k\| \le \tau t_k L_H \|d^k\|, \quad \forall \tau \in [0, 1], \quad \forall k \ge 0. \tag{3.4.1}$$

Moreover, since $\Omega + B(0, b_1)$ is compact and f is twice continuously differentiable, ∇f is Lipschitz continuous on $\Omega + B(0, b_1)$. Thus, there exists a positive constant L_g such that

$$\|\nabla^2 f(x)\| \le L_g, \quad \forall x \in \Omega + B(0, b_1). \tag{3.4.2}$$

Now, we show that the condition (3.2.1) in Armijo's step size rule holds if the step size t_k is less than a specific value independent of k.

Lemma 3.4.2. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Suppose also that*

$$t_k \le \sqrt{\frac{4(1-\alpha)c_2^2}{L_H L_f^{1-2\delta}}}.$$

Then,

$$f_k - f(x^k + t_k d^k) \ge -\alpha t_k {g^k}^T d^k.$$

Proof. In a way similar to the proof of Lemma 3.3.3, we have

$$\begin{aligned} & f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \\ & = (1-\alpha) t_k {d^k}^T (H_k + c_1 \Lambda_k I + c_2 \|g^k\|^\delta I) d^k - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau. \end{aligned}$$

It then follows from Assumption 3.4.1 (b), the positive definiteness of $H_k + c_1 \Lambda_k I$ that

$$\begin{aligned} & f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \\ & \ge \frac{1}{2} t_k {d^k}^T (H_k + c_1 \Lambda_k I) d^k + (1-\alpha) t_k c_2 \|g^k\|^\delta \|d^k\|^2 - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau \\ & \ge \frac{1}{2} t_k^2 {d^k}^T H_k d^k + (1-\alpha) t_k c_2 \|g^k\|^\delta \|d^k\|^2 - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau \\ & = \frac{1}{2} t_k {d^k}^T \left(2(1-\alpha) c_2 \|g^k\|^\delta I - t_k \int_0^1 (\nabla^2 f(x^k + \tau t_k d^k) - H_k) \mathrm{d}\tau \right) d^k. \end{aligned} \tag{3.4.3}$$

where the second inequality follows from the fact that $t_k \le 1$. From (3.4.1), we have

$$\begin{aligned} \int_0^1 {d^k}^T (\nabla^2 f(x^k + \tau t_k d^k) - H_k) d^k \mathrm{d}\tau & \le \|d^k\|^2 \int_0^1 \|\nabla^2 f(x^k + \tau t_k d^k) - H_k\| \mathrm{d}\tau \\ & \le t_k L_H \|d^k\|^3 \int_0^1 \tau \mathrm{d}\tau \\ & = \frac{1}{2} t_k L_H \|d^k\|^3. \end{aligned}$$

It then follows from (3.4.3) that

$$\begin{aligned} f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k & \ge \frac{1}{4} t_k \|d^k\|^2 (4(1-\alpha) c_2 \|g^k\|^\delta - t_k^2 L_H \|d^k\|) \\ & \ge \frac{t_k L_H \|g^k\|^{1-\delta} \|d^k\|^2}{4 c_2} \left(\frac{4(1-\alpha) c_2^2}{L_H \|g^k\|^{1-2\delta}} - t_k^2 \right) \\ & \ge \frac{t_k L_H \|g^k\|^{1-\delta} \|d^k\|^2}{4 c_2} \left(\frac{4(1-\alpha) c_2^2}{L_H L_f^{1-2\delta}} - t_k^2 \right) \\ & \ge 0, \end{aligned}$$

where the second inequality follows from Lemma 3.3.1, the third inequality follows from (3.3.2), and the last inequality follows from the assumption on t_k. □

From Lemma 3.3.3, we show that the step size t_k determined in Step 4 of the extended RNM is bounded below by some positive constant independent of k.

Lemma 3.4.3. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then,*

$$t_k \geq t_{\min},$$

where

$$t_{\min} := \min\left(1, \sqrt{\frac{4(1-\alpha)\beta^2 c_2^2}{L_H L_f^{1-2\delta}}}\right).$$

Proof. From Lemma 3.4.2 and the definition of t_k in Step 4 of the extended RNM, t_k must be

$$t_k \geq \min\left(1, \sqrt{\frac{4(1-\alpha)\beta^2 c_2^2}{L_H L_f^{1-2\delta}}}\right).$$

Otherwise βt_k satisfies Armijo's rule, which contradicts the definition of t_k. □

From Lemma 3.4.3, we show that the number of backtracking steps is bounded from above by some positive constant independent of k.

Theorem 3.4.1. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then,*

$$l_k \leq l_{\max},$$

where

$$l_{\max} := \frac{\ln t_{\min}}{\ln \beta}.$$

Proof. From Armijo's rule and Lemma 3.4.3, we have $l_k \ln \beta \geq \ln t_{\min}$. Since $\ln \beta < 0$, it follows that $l_k \leq \frac{\ln t_{\min}}{\ln \beta}$. □

Remark 3.4.1. *Since Polyak's RNM* [57] *uses a special step size* (3.1.2)*, his method does not need a backtracking scheme. However, the step size of Polyak's RNM cannot be used when the Lipschitz constant L_f is unknown.*

Next, we give a lower bound of the reduction $f_k - f_{k+1}$ when $k < K_{\text{outer}}$.

Lemma 3.4.4. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then, for all k such that $k < K_{\text{outer}}$,*

$$f_k - f_{k+1} > p\varepsilon_g^2,$$

where

$$p := \frac{\alpha t_{\min}}{(1+c_1)L_g + c_2 L_f^{\delta}}.$$

Proof. For all k such that $k < K_{\text{outer}}$, we have $\|g^k\| > \varepsilon_g$. Thus, from Lemma 3.4.3 and (3.4.2), we can get the desired inequality in a way similar to the proof of Lemma 3.3.5. □

Now, we give an upper bound of K_{outer}.

Theorem 3.4.2. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then,*

$$K_{\text{outer}} \le \left\lceil \frac{f_0 - f_{\min}}{p} \varepsilon_g^{-2} + 1 \right\rceil .$$

Proof. Let K be $\lceil ((f_0 - f_{\min})\varepsilon_g^{-2}/p) + 1 \rceil$. Suppose the contrary, i.e., $K_{\text{outer}} > K$. It then follows from (3.3.1) and Lemma 3.4.4 that

$$f_0 - f_{\min} \ge f_0 - f_K \ge \sum_{j=0}^{K-1} (f_j - f_{j+1}) > \sum_{j=0}^{K-1} p\varepsilon_g^2 = p\varepsilon_g^2 K. \tag{3.4.4}$$

On the other hand, we have

$$p\varepsilon_g^2 K = p\varepsilon_g^2 \left\lceil \frac{f_0 - f_{\min}}{p} \varepsilon_g^{-2} + 1 \right\rceil > f_0 - f_{\min}$$

from the definition of K. This contradicts (3.4.4), and hence we obtain the result of the theorem. □

By using Theorem 3.4.2, we show the global complexity bound K_{total} of the extended RNM.

Theorem 3.4.3. *Suppose that Assumptions* 3.3.1 *and* 3.4.1 *hold. Then,*

$$K_{\text{total}} \le (l_{\max} + 1) K_{\text{outer}},$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$.

Proof. From the definition of K_{total}, we have

$$\begin{aligned} K_{\text{total}} &= \sum_{k=0}^{K_{\text{outer}}-1} (l_k + 1) \le \sum_{k=0}^{K_{\text{outer}}-1} (l_{\max} + 1) \\ &= (l_{\max} + 1) K_{\text{outer}} \le (l_{\max} + 1) \left\lceil \frac{f_0 - f_{\min}}{p} \varepsilon_g^{-2} + 1 \right\rceil , \end{aligned}$$

where the first inequality follows from Theorem 3.4.1, and the last inequality follows from Theorem 3.4.2. Thus, we have $K = O(\varepsilon_g^{-2})$. □

Remark 3.4.2. *Polyak shows that* $f_k - f_{k+1} \ge \hat{p} \|g^k\|^4$ *for his method in the proof of* [57, Theorem 3]*, where* $\hat{p}$ *is some positive constant. Thus, the global complexity bound of his method is* $O(\varepsilon_g^{-4})$*, which is worse than the global complexity bound* $O(\varepsilon_g^{-2})$ *given in Theorem* 3.4.3*. Note, however, that the proof of* [57, Theorem 3] *does not use the Lipschitz continuity of* $\nabla^2 f$*. Moreover, if* f *is strongly convex, by modifying the proof of* [57, Theorem 3]*, we can show that Polyak's method also has the global complexity bound* $O(\varepsilon_g^{-2})$ *without the Lipschitz continuity of* $\nabla^2 f$*.*

3.5 Local convergence

In this section, we show that the extended RNM has a superlinear convergence under the local error bound condition. In order to prove the superlinear convergence, we use techniques similar to

[68] and [20]. In [68], Yamashita and Fukushima showed that the Levenberg-Marquardt method has a quadratic rate of convergence under the local error bound condition. Similarly, in [20], Dan, Yamashita and Fukushima showed that the inexact Levenberg-Marquardt method has a superlinear rate of convergence under the local error bound condition.

First, we make the following assumptions.

Assumption 3.5.1.

(a) $0 < \delta < 1$.

(b) $\alpha \leq 1/2$.

(c) *There exists a local optimal solution* x^* *of the problem* (3.1.1).

(d) $\nabla^2 f$ *is locally Lipschitz continuous near* x^**, i.e., there exist constants* $b_2 \in (0, 1)$ *and* $\bar{L}_H > 0$ *such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \bar{L}_H \|x - y\|, \quad \forall x, y \in B(x^*, b_2).$$

(e) $\|\nabla f(x)\|$ *provides a local error bound for the problem* (3.1.1) *on* $B(x^*, b_2)$*, i.e., there exists a constant* $\kappa_1 > 0$ *such that*

$$\kappa_1 \mathrm{dist}(x, X^*) \leq \|\nabla f(x)\|, \quad \forall x \in B(x^*, b_2),$$

where X^* *is the local optimal solution set of* (3.1.1).

Note that, under Assumption 3.5.1, the following inequality holds.

$$\|\nabla f(x) - \nabla f(y) - \nabla f(x)(x - y)\| \leq \frac{1}{2}\bar{L}_H \|x - y\|^2, \quad \forall x, y \in B(x^*, b_2). \tag{3.5.1}$$

Since f is twice continuously differentiable, there exist a positive constant $\bar{L}_g$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq \bar{L}_g \|x - y\|, \quad \forall x, y \in B(x^*, b_2). \tag{3.5.2}$$

In what follows, $\bar{x}^k$ denotes an arbitrary vector such that

$$\|x^k - \bar{x}^k\| = \mathrm{dist}(x^k, X^*), \quad \bar{x}^k \in X^*.$$

In the case where f is convex, Li, Fukushima, Qi and Yamashita [38] showed that the RNM has a quadratic rate of convergence under the local error bound condition. The convexity of f implies $\Lambda_k \equiv 0$. However, since f is not necessarily convex, it is not always true that $\Lambda_k = 0$. Therefore, we now investigate the relationship between Λ_k and $\mathrm{dist}(x^k, X^*)$. To this end, we need the following property on a singular matrix.

Lemma 3.5.1. *Suppose that* $M \in \mathbb{R}^{n \times n}$ *is singular, then* $\|I - M\| \geq 1$.

Proof. It directly follows from [31, Corollary 5.6.16]. □

By using Lemma 3.5.1, we show the following key lemma for superlinear convergence.

Lemma 3.5.2. *Suppose that Assumption* 3.5.1 *holds. Suppose also that* $x^k \in B(x^*, b_2/2)$*. Then*

$$\Lambda_k \leq \bar{L}_H \mathrm{dist}(x^k, X^*).$$

Proof. When $H_k \succeq 0$, we have $\Lambda_k = 0$. Thus, the desired inequality holds. Next, we assume $\lambda_{\min}(H_k) < 0$. Let $\bar{\lambda}_k^{(l)}$ be the l-th largest eigenvalue of $\nabla^2 f(\bar{x}^k)$. Since $\bar{x}^k \in X^*$, we have $\bar{\lambda}_k^{(l)} \geq 0$. Moreover, since $\nabla^2 f(\bar{x}^k)$ is a real symmetric matrix, $\nabla^2 f(\bar{x}^k)$ can be diagonalized by some orthogonal matrix $\bar{Q}_k$, i.e.,

$$\bar{Q}_k^T \nabla^2 f(\bar{x}^k) \bar{Q}_k = \mathrm{diag}(\bar{\lambda}_k^{(l)}),$$

where $\mathrm{diag}(\bar{\lambda}_k^{(l)})$ denotes the diagonal matrix whose (l, l) element is $\bar{\lambda}_k^{(l)}$. Then, we have

$$\begin{aligned} \lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k &= \lambda_{\min}(H_k)I - \bar{Q}_k^T \Big(\nabla^2 f(\bar{x}^k) + (H_k - \nabla^2 f(\bar{x}^k))\Big)\bar{Q}_k \\ &= \lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k. \end{aligned}$$

Since $\bar{Q}_k^T H_k \bar{Q}_k$ has the eigenvalue $\lambda_{\min}(H_k)$, $\lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k$ is singular, and hence $\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k$ is also singular. On the other hand, $\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})$ is nonsingular because $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(l)} \geq 0$.

Now, we denote M as

$$M := \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k\Big).$$

Then, M is singular. It then follows from Lemma 3.5.1 that

$$\begin{aligned} 1 &\leq \|I - M\| \\ &= \left\| I - \left(I - \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k\right)\right\| \\ &= \left\| \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k \right\| \\ &\leq \left\| \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \right\| \cdot \|\bar{Q}_k^T (H_k - \nabla^2 f(\bar{x}^k))\bar{Q}_k\| \\ &= \left\| \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \right\| \cdot \|H_k - \nabla^2 f(\bar{x}^k)\|. \end{aligned} \tag{3.5.3}$$

We consider $\|(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}))^{-1}\|$ and $\|H_k - \nabla^2 f(\bar{x}^k)\|$ separately. Since $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(l)} \geq 0$, we have

$$\begin{aligned} \left\| \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \right\| &= \max_{1 \leq l \leq n} \left|\lambda_{\min}(H_k) - \bar{\lambda}_k^{(l)}\right|^{-1} = \frac{1}{\min\limits_{1 \leq l \leq n} \left|\lambda_{\min}(H_k) - \bar{\lambda}_k^{(l)}\right|} \\ &\leq \frac{1}{|\lambda_{\min}(H_k)|} = \frac{1}{\Lambda_k}, \end{aligned} \tag{3.5.4}$$

where the last equality follows from the definition of Λ_k. Next, we consider $\|H_k - \nabla^2 f(\bar{x}^k)\|$. Since $x^k \in B(x^*, b_2/2)$, we have

$$\|\bar{x}^k - x^*\| \leq \|\bar{x}^k - x^k\| + \|x^k - x^*\| \leq \|x^* - x^k\| + \|x^k - x^*\| \leq b_2,$$

and hence $\bar{x}^k \in B(x^*, b_2)$. It then follows from Assumption 3.5.1 (d) that

$$\|H_k - \nabla^2 f(\bar{x}^k)\| \leq \bar{L}_H \|x^k - \bar{x}^k\| = \bar{L}_H \mathrm{dist}(x^k, X^*). \tag{3.5.5}$$

Therefore, we have from (3.5.3) – (3.5.5) that

$$1 \leq \frac{\bar{L}_H \mathrm{dist}(x^k, X^*)}{\Lambda_k},$$

which is the desired inequality. □

From this lemma, we can show the superlinear convergence in a way similar to [68] and [20]. First, we show that $\|d^k\| = O(\mathrm{dist}(x^k, X^*))$

Lemma 3.5.3. *Suppose that Assumption* 3.5.1 *holds. Suppose also that* $x^k \in B(x^*, b_2/2)$*. Then,*

$$\|d^k\| \leq \kappa_2 \mathrm{dist}(x^k, X^*),$$

where

$$\kappa_2 := \frac{\bar{L}_H}{2c_2\kappa_1^\delta} + \max\left(1, \frac{1}{c_1 - 1}\right).$$

Proof. First note that $\nabla f(\bar{x}^k) = 0$. Since $x^k \in B(x^*, b_2/2)$, we have $\bar{x}^k \in B(x^*, b_2)$. It then follows from the definition of d^k and (3.5.1) that

$$\begin{aligned}
\|d^k\| &= \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} g^k\right\| \\
&= \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\left(g^k - \nabla f(\bar{x}^k) - H_k(x^k - \bar{x}^k) + H_k(x^k - \bar{x}^k)\right)\right\| \\
&\leq \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\right\| \cdot \|g^k - \nabla f(\bar{x}^k) - H_k(x^k - \bar{x}^k)\| \\
&\quad + \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k\right\| \cdot \|x^k - \bar{x}^k\| \\
&\leq \frac{1}{2}\bar{L}_H \|x^k - \bar{x}^k\|^2 \cdot \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\right\| \\
&\quad + \|x^k - \bar{x}^k\| \cdot \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k\right\| \\
&\leq \frac{1}{2}\bar{L}_H \mathrm{dist}(x^k, X^*)^2 \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\right\| \cdot \|g^k - \nabla f(\bar{x}^k) - H_k(x^k - \bar{x}^k)\| \\
&\quad + \mathrm{dist}(x^k, X^*) \left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k\right\|.
\end{aligned} \tag{3.5.6}$$

From the positive semidefiniteness of $H_k + c_1\Lambda_k$ and Assumption 3.5.1 (e), we have

$$\begin{aligned}\left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\right\| &= \lambda_{\max}\Big((H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1}\Big) \\ &= \frac{1}{\lambda_{\min}(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)} \\ &\le \frac{1}{c_2\|g^k\|^\delta} \\ &\le \frac{1}{c_2\kappa_1^\delta \operatorname{dist}(x^k, X^*)^\delta}.\end{aligned} \tag{3.5.7}$$

Next, we consider $\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k\|$. Let $\lambda_k^{(l)}$ be the l-th largest eigenvalue of H_k. Then, the eigenvalues of $(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k$ are given by

$$\frac{\lambda_k^{(l)}}{\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g^k\|^\delta}, \quad 1 \le l \le n.$$

Now, we consider two cases: (i) $\lambda_k^{(l)} \ge 0$ and (ii) $\lambda_k^{(l)} < 0$.

Case (i): This case implies that

$$\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g^k\|^\delta\right|} \le 1.$$

Case (ii): In this case, since $-\Lambda_k = \lambda_{\min}(H_k) \le \lambda_k^{(l)} < 0$, we have $\lambda_k^{(l)} - \lambda_{\min}(H_k) \ge 0$ and $|\lambda_k^{(l)}| \le |\lambda_{\min}(H_k)|$. Therefore, we have

$$\begin{aligned}\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g^k\|^\delta\right|} &= \frac{\left|\lambda_k^{(l)}\right|}{\left|(\lambda_k^{(l)} - \lambda_{\min}(H_k)) - (c_1 - 1)\lambda_{\min}(H_k) + c_2\|g^k\|^\delta\right|} \\ &\le \frac{|\lambda_{\min}(H_k)|}{\lambda_k^{(l)} - \lambda_{\min}(H_k) + (c_1 - 1)\,|\lambda_{\min}(H_k)| + c_2\|g^k\|^\delta} \\ &\le \frac{1}{c_1 - 1}.\end{aligned}$$

Thus, we have

$$\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g^k\|^\delta\right|} \le \max\left(1, \frac{1}{c_1 - 1}\right), \quad 1 \le l \le n,$$

and hence

$$\left\|(H_k + c_1\Lambda_k I + c_2\|g^k\|^\delta I)^{-1} H_k\right\| \le \max\left(1, \frac{1}{c_1 - 1}\right). \tag{3.5.8}$$

From (3.5.6) – (3.5.8), we have

$$\begin{aligned}\|d^k\| &\le \frac{\bar{L}_H}{2c_2\kappa_1^\delta}\mathrm{dist}(x^k, X^*)^{2-\delta} + \max\left(1, \frac{1}{c_1-1}\right)\mathrm{dist}(x^k, X^*)\\ &\le \left(\frac{\bar{L}_H}{2c_2\kappa_1^\delta} + \max\left(1, \frac{1}{c_1-1}\right)\right)\mathrm{dist}(x^k, X^*),\end{aligned}$$

where the last inequality follows from the facts that $\mathrm{dist}(x^k, X^*) \le \|x^k - x^*\| \le b_2/2 < 1$ and $\delta < 1$. □

From the above lemma, we can show that $x^k + \tau t_k d^k \in B(x^*, b_2)$ for any $\tau \in [0,1]$ if x^k is sufficiently close to x^*.

Lemma 3.5.4. *Suppose that Assumption* 3.5.1 *holds. Let* $b_3 := b_2/(\kappa_2+1)$. *Suppose also that* $x^k \in B(x^*, b_3)$. *Then*

$$x^k + \tau t_k d^k \in B(x^*, b_2), \quad \forall \tau \in [0,1].$$

Proof. Since $b_3 \le b_2/2$, we have $x^k \in B(x^*, b_2/2)$. Thus, from Lemma 3.5.3, we have

$$\begin{aligned}\|x^k + \tau t_k d^k - x^*\| &\le \|x^k - x^*\| + \|d^k\| \le \|x^k - x^*\| + \kappa_2 \mathrm{dist}(x^k, X^*)\\ &\le \|x^k - x^*\| + \kappa_2\|x^k - x^*\| \le (\kappa_2+1)b_3 = b_2,\end{aligned}$$

which is the desired inequality. □

Now, we show that $l_k = 0$ (that is, $t_k = 1$) is accepted in Step 3 if x^k is sufficiently close to x^*.

Lemma 3.5.5. *Suppose that Assumption* 3.5.1 *holds. Let*

$$b_4 := \min\left(b_3, \left(\frac{2(1-\alpha)c_2\kappa_1^\delta}{\kappa_2\bar{L}_H}\right)^{\frac{1}{1-\delta}}\right).$$

Suppose that $x^k \in B(x^*, b_4)$. *Then,*

$$t_k = 1.$$

Proof. In a way similar to the proof of Lemma 3.4.2, we have from Lemma 3.5.4 and Assumption 3.5.1 (d) that

$$f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \ge \frac{1}{4}t_k\|d^k\|^2(4(1-\alpha)c_2\|g^k\|^\delta - t_k^2\bar{L}_H\|d^k\|).$$

It then follows from Lemma 3.5.3 and Assumption 3.5.1 (e) that

$$\begin{aligned}f(x^k) - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k &\ge \frac{t_k\bar{L}_H}{4}\left(\frac{4(1-\alpha)c_2\kappa_1^\delta}{\kappa_2\bar{L}_H\mathrm{dist}(x^k, X^*)^{1-\delta}} - t_k^2\right)\|d^k\|^3\\ &\ge \frac{t_k\bar{L}_H}{4}\left(\frac{4(1-\alpha)c_2\kappa_1^\delta}{\kappa_2\bar{L}_H\|x^k - x^*\|^{1-\delta}} - t_k^2\right)\|d^k\|^3\\ &\ge \frac{t_k\bar{L}_H}{4}(1-t_k^2)\|d^k\|^3,\end{aligned}$$

where the last inequality follows from $x^k \in B(x^*, b_4)$. Thus, we have $t_k = 1$. □

Next, we show that $\mathrm{dist}(x^k, X^*)$ converges to 0 superlinearly, when $\{x^k\}$ lie in a neighborhood of x^*.

Lemma 3.5.6. *Suppose that Assumption* 3.5.1 *holds. Suppose also that* $x^k \in B(x^*, b_4)$. *Then,*

$$\mathrm{dist}(x^{k+1}, X^*) = O\left(\mathrm{dist}(x^k, X^*)^{1+\delta}\right).$$

Proof. From Lemma 3.5.5, we have $t_k = 1$, and hence $x^{k+1} = x^k + d^k$. It then follows from Lemma 3.5.4 that $x^{k+1} \in B(x^*, b_2)$. Thus, from Assumption 3.5.1 (e), we have

$$\begin{aligned}
\mathrm{dist}(x^{k+1}, X^*) &\le \frac{1}{\kappa_1}\|g^{k+1}\| \\
&\le \frac{1}{\kappa_1}\|H_k d^k + g^k\| + \frac{\bar{L}_H}{2\kappa_1}\|d^k\|^2 \\
&= \frac{1}{\kappa_1}\left\|c_1 \Lambda_k d^k + c_2\|g^k\|^\delta d^k\right\| + \frac{\bar{L}_H}{2\kappa_1}\|d^k\|^2 \\
&\le \frac{c_1\Lambda_k}{\kappa_1}\|d^{k*}\| + \frac{c_2}{\kappa_1}\|g^k\|^\delta\|d^k\| + \frac{\bar{L}_H}{2\kappa_1}\|d^k\|^2, \qquad (3.5.9)
\end{aligned}$$

where the second inequality follows from (3.5.1), and the first equality follows from the definition of d^k. Since $x^k \in B(x^*, b_2/2)$, we have $\bar{x}^k \in B(x^*, b_2)$. Then, from (3.5.2) and the fact that $\nabla f(\bar{x}^k) = 0$, we have

$$\|g^k\|^\delta = \|g^k - \nabla f(\bar{x}^k)\|^\delta \le \bar{L}_g^\delta \mathrm{dist}(x^k, X^*)^\delta. \qquad (3.5.10)$$

Therefore, from (3.5.9), (3.5.10) and Lemmas 3.5.2 and 3.5.3, we have

$$\begin{aligned}
\mathrm{dist}(x^{k+1}, X^*) &\le \frac{c_1\kappa_2\bar{L}_H}{\kappa_1}\mathrm{dist}(x^k, X^*)^2 + \frac{c_2\kappa_2\bar{L}_g^\delta}{\kappa_1}\mathrm{dist}(x^k, X^*)^{1+\delta} + \frac{\kappa_2^2\bar{L}_H}{2\kappa_1}\mathrm{dist}(x^k, X^*)^2 \\
&\le \frac{\kappa_2(2c_1\bar{L}_H + 2c_2\bar{L}_g^\delta + \kappa_2\bar{L}_H)}{2\kappa_1}\mathrm{dist}(x^k, X^*)^{1+\delta}.
\end{aligned}$$

This completes the proof. □

From Lemma 3.5.6, there exists a positive constant $b_5 \le b_4$ such that

$$\mathrm{dist}(x^k, X^*) \le b_5 \Rightarrow \mathrm{dist}(x^{k+1}, X^*) \le \frac{1}{2}\mathrm{dist}(x^k, X^*).$$

Now, we give a sufficient condition for $x^k \in B(x^*, b_4)$ for all k.

Lemma 3.5.7. *Suppose that Assumption* 3.5.1 *holds. Let* $b_6 := \frac{1}{1+2\kappa_2}b_5$. *Suppose that* $x^0 \in B(x^*, b_6)$. *Then,* $x^k \in B(x^*, b_5)$ *for all* k.

Proof. In a way similar to the proof of [68, Lemma 2.3], we obtain the desired result. □

By using Lemmas 3.5.6 and 3.5.7, we give the rate of convergence of the extended RNM.

Theorem 3.5.1. *Suppose that Assumption* 3.5.1 *holds. Suppose also that* $x^0 \in B(x^*, b_6)$. *Then,* $\{\mathrm{dist}(x^k, X^*)\}$ *converges to* 0 *at the rate of* $1 + \delta$. *Moreover,* $\{x^k\}$ *converges to a local optimal solution* $\hat{x} \in B(x^*, b_5)$.

Proof. The first part of the theorem directly follows from Lemmas 3.5.6 and 3.5.7. Moreover, in a way similar to the proof of [68, Theorem 2.1], we obtain the result of the second part of the theorem. □

Remark 3.5.1. *Note that in a way similar to the proof of* [38, Theorem 3.2], *we can prove that* $\{x^k\}$ *converges to* $\hat{x}$ *at the rate of* $1+\delta$.

We get a rapid local convergence if we take a larger δ. However, we cannot guarantee the quadratic convergence since Lemma 3.5.5 requires $\delta < 1$. When f is strongly convex and $\delta = 1$, we can show the following lemma which is an alternate of Lemma 3.5.5.

Lemma 3.5.8. *Suppose that Assumptions* 3.5.1 (c) *and* (d) *hold. Suppose also that* $\delta = 1$, $\alpha < 1/2$, *and that* f *is strongly convex. Then,* $t_k = 1$ *for sufficiently large* k.

Proof. In a way similar to the proof of Lemma 3.3.3, we have

$$\begin{aligned} & f_k - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \\ &= (1-\alpha) t_k {d^k}^T (H_k + c_1 \Lambda_k I + c_2 \|g^k\| I) d^k - \frac{1}{2} t_k^2 \int_0^1 {d^k}^T \nabla^2 f(x^k + \tau t_k d^k) d^k \mathrm{d}\tau. \end{aligned}$$

Since f is strongly convex, $\Lambda_k = 0$ for all k. Moreover, there exists $\sigma > 0$ such that

$$v^T \nabla^2 f(x) v \geq \sigma \|v\|^2, \quad \forall x, v \in \mathbb{R}^n. \tag{3.5.11}$$

Thus, we have

$$\begin{aligned} & f(x^k) - f(x^k + t_k d^k) + \alpha t_k {g^k}^T d^k \\ &= \left(\frac{1}{2} - \alpha\right) t_k {d^k}^T H_k d^k + (1-\alpha) c_2 t_k \|g^k\| \cdot \|d^k\|^2 + \frac{1}{2} t_k \int_0^1 {d^k}^T (H_k - t_k \nabla^2 f(x^k + \tau t_k d^k)) d^k \mathrm{d}\tau \\ &\geq \left(\frac{1}{2} - \alpha\right) t_k {d^k}^T H_k d^k + (1-\alpha) c_2 t_k \|g^k\| \cdot \|d^k\|^2 + \frac{1}{2} t_k^2 \int_0^1 {d^k}^T (H_k - \nabla^2 f(x^k + \tau t_k d^k)) d^k \mathrm{d}\tau \\ &\geq \left(\frac{1}{2} - \alpha\right) t_k {d^k}^T H_k d^k + (1-\alpha) c_2 t_k \|g^k\| \cdot \|d^k\|^2 - \frac{1}{2} t_k^2 \|d^k\|^2 \int_0^1 \|H_k - \nabla^2 f(x^k + \tau t_k d^k)\| \mathrm{d}\tau \\ &\geq \left(\frac{1}{2} - \alpha\right) t_k \sigma \|d^k\|^2 + (1-\alpha) c_2 t_k \|g^k\| \cdot \|d^k\|^2 - \frac{1}{4} t_k^3 \bar{L}_H \|d^k\|^3 \\ &= t_k \|d^k\|^2 \left(\left(\frac{1}{2} - \alpha\right) \sigma + (1-\alpha) c_2 \|g^k\| - \frac{1}{4} t_k^2 \bar{L}_H \|d^k\| \right), \end{aligned} \tag{3.5.12}$$

where the first inequality follows from the positive definiteness of H_k and the fact that $t_k \leq 1$. and the third inequality follows from (3.5.8) and Assumption 3.5.1 (d). Since f is strongly convex, Assumption 3.5.1 (e) holds for some κ_1. Thus, Lemma 3.5.3 holds with $\delta = 1$. Moreover, since $x^k \to x^*$ from Theorem 3.3.1, Lemma 3.5.3 implies $\|d^k\| \to 0$. It then follow from (3.5.12) that $t_k = 1$ for sufficiently large k. □

Using this lemma, we can show the quadratic rate of convergence of the extended RNM in a way similar to Theorem 3.5.1.

Theorem 3.5.2. *Suppose that the assumptions of Lemma* 3.5.8 *hold. Then,* $\{x^k\}$ *converges to the unique solution* x^* *quadratically.*

3.6 Concluding remarks

In this chapter, we have proposed an RNM extended to the unconstrained nonconvex minimization problem. We have shown that the extended RNM has a global convergence and a superlinear convergence under appropriate conditions. Moreover, we have shown that the global complexity bound of the extended RNM is $O(\varepsilon_g^{-2})$ when $\nabla^2 f$ is Lipschitz continuous. The obtained global complexity bound of the extended RNM is better than the global complexity bound of Polyak's RNM.

If the parameter c_2 in the extended RNM is large enough, then the constant $t_{\min}$ given in Lemma 3.4.3 is always 1. Thus, the constant p given in Lemma 3.4.4 gets very small. In this case, the concrete global complexity bound indecated in Theorem 3.4.2 becomes very large. Therefore, the choice of the parameters δ, c_1 and c_2 is very important in the implementation of the extended RNM.

From Theorem 3.5.1, we can get a rapid local convergence as we take a larger δ. On the other hand, in Section 3.4, we assumed $\delta \leq 1/2$ to give the global complexity bound. It is an interesting subject to investigate the tightness of δ.

Chapter 4

A regularized Newton method without line search for the unconstrained minimization problem and its convergence properties

4.1 Introduction

As is the case in Chapter 3, we consider the regularized Newton method (RNM) for the following unconstrained minimization problem.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x), \tag{4.1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function.

For the k-th iterative point x^k and a positive parameter μ_k, a search direction $d^k(\mu_k)$ of the RNM is given by

$$d^k(\mu_k) := -(H_k + \mu_k I)^{-1} g^k.$$

In Chapter 3, we extended the RNM to the unconstrained nonconvex minimization problem, and showed its convergence properties. The extended RNM adopts the regularization parameter μ_k as

$$\mu_k := c_1 \Lambda_k + c_2 \|g^k\|^\delta,$$

where c_1, c_2, δ are given constants such that $c_1 > 1, c_2 > 0, \delta \geq 0$, and Λ_k is defined by

$$\Lambda_k := \max(0, -\lambda_{\min}(H_k)).$$

For global convergence, the extended RNM uses line search. Quite recently, Shen, Chen and Liang [60] proposed an RNM for the unconstrained nonconvex minimization problem. In their RNM, a search direction $d^k(\mu_k)$ is given by

$$d^k(\mu_k) := -(H_k + E_k(\mu_k))^{-1} g^k,$$

where $E_k(\mu_k)$ is a regularization matrix. The matrix $E_k(\mu_k)$ is calculated by the modified Cholesky factorization [15]. Thus, their RNM does not necessarily compute the exact minimum eigenvalue $\lambda_{\min}(H_k)$. Their RNM also uses line search for global convergence.

As described in Section 1.2, the trust-region Newton method (trust-region NM) [17] is one of the solution methods for (4.1.1). A search direction of the trust-region NM is given as a solution of the subproblem (1.2.2). For global convergence, the trust-region NM adaptively controls the trust-region radius instead of using a step size at each iteration. From various numerical experiments, it is known that the number of function evaluations of the trust-region NM required to get a solution tends to be less as compared to that of the Newton-type methods with some line search method. However, since we have to solve the nonlinear equation (1.2.3), a lot of computational time may be required to get the search direction of the trust-region NM. Therefore, it is desirable to construct a solution method whose behavior is similar to the trust-region NM, and subproblems can be solved as the RNM.

In this chapter, we proposed an RNM without line search. In order to guarantee the global convergence, it controls the regularization parameter as the trust-region NM. The proposed RNM only solves linear equations to get a search direction. We show that the proposed RNM has global convergence, and superlinear convergence under the local error bound condition. We also give global complexity bounds of the proposed RNM. In particular, we show that the global complexity bounds are $O(\varepsilon_g^{-2})$ when f is nonconvex, $O(\varepsilon_g^{-\frac{5}{3}})$ and $O(\varepsilon_f^{-\frac{2}{3}})$ when f is convex, and $O(\varepsilon_g^{-1})$ and $O(\log \varepsilon_f^{-1})$ when f is strongly convex.

This chapter is organized as follows. In the next section, we propose an RNM that controls the regularization parameter at each iteration. In Section 4.3, we show its global convergence. In Section 4.4, we give its global complexity bounds. In Section 4.5, we establish superlinear convergence under the local error bound condition. Then, numerical results are presented and discussed in Section 4.6. Finally Section 4.7 concludes this chapter.

4.2 Proposed algorithm

In this section, we propose an RNM without line search. For a given positive parameter ν_k, we compute a search direction $d^k(\nu_k)$ defined by

$$d^k(\nu_k) := -(H_k + E_k(\nu_k))^{-1} g^k,$$

where $E_k(\nu_k)$ is a regularization matrix such that $H_k + E_k(\nu_k)$ is positive definite. To guarantee good convergence properties, we assume the following properties for $E_k(\nu)$.

Assumption 4.2.1.
Let δ be a constant such that $\delta \geq 0$.

(a) $E_k(\nu)$ *is symmetric and positive semidefinite.*

(b) *There exist positive constants c_1 and c_2 such that*

$$\lambda_{\min}(H_k + E_k(\nu)) \geq c_1 \Lambda_k + c_2 \nu \|g^k\|^\delta.$$

(c) *There exist positive constants* c_3 *and* c_4 *such that*

$$\|E_k(\nu)\| \leq c_3\Lambda_k + c_4\nu\|g^k\|^\delta.$$

When calculating Λ_k is not expensive, we can set $E_k(\nu)$ as $c\Lambda_k I + \nu\|g^k\|^\delta I$ with $c > 1$. In particular, when f is convex, $\Lambda_k = 0$. Then, we can simply set $E_k(\nu)$ as $\nu\|g^k\|^\delta I$. On the other hand, when calculating Λ_k is expensive, we may adopt the modified Cholesky factorization [15] which can implicitly compute $E_k(\nu)$ satisfying Assumption 4.2.1. We illustrate the brief idea as follows.

First, by using the bounded Bunch-Kaufman pivoting strategy [2], we factorize H_k as

$$P_k H_k P_k^T = Q_k D_k Q_k^T, \tag{4.2.1}$$

where P_k is a permutation matrix, Q_k is a unit lower triangular matrix, and D_k is a block diagonal matrix with diagonal blocks of dimension 1 or 2. Then, we provide $H_k + E_k(\nu)$ as

$$H_k + E_k(\nu) := P_k^T Q_k \Big(D_k + (c\max(0, -\lambda_{\min}(D_k)) + \nu\|g^k\|^\delta)I\Big) Q_k^T P_k, \tag{4.2.2}$$

where c is a positive constant such that $c > 1$. Note that, since D_k is a block diagonal matrix with diagonal blocks of dimension 1 or 2, we can easily compute $\lambda_{\min}(D_k)$.

Based on techniques used in [15], we show that $E_k(\nu)$ satisfies Assumption 4.2.1. To this end, we introduce the following lemmas [15, 31].

Lemma 4.2.1. *Let* $M \in \mathbb{R}^{n\times n}$ *be a nonsingular matrix,* $S \in \mathbb{R}^{n\times n}$ *be a symmetric matrix,* $\lambda_i(\cdot)$ *be the* i*-th largest eigenvalue of a matrix. Then,*

$$\lambda_i(MSM^T) = \theta_i\lambda_i(S),$$

where $\lambda_{\min}(MM^T) \leq \theta_i \leq \lambda_{\max}(MM^T)$.

Lemma 4.2.2. *For the unit lower triangular matrix* Q_k *given by the bounded Bunch-Kaufman pivoting strategy, it satisfies that*

$$\lambda_{\max}(Q_k Q_k^T) \leq 4n^2 - 3n,$$
$$\lambda_{\min}(Q_k Q_k^T) \geq (3.781)^{2-2n}.$$

By using Lemmas 4.2.1 and 4.2.2, we give the following theorem.

Theorem 4.2.1. *The matrix* $E_k(\nu)$ *defined in* (4.2.2) *holds Assumption* 4.2.1.

Proof. First, from (4.2.1) and (4.2.2), we have

$$E_k(\nu) = (c\max(0, -\lambda_{\min}(D_k)) + \nu\|g^k\|^\delta)P_k^T Q_k Q_k^T P_k, \tag{4.2.3}$$

and hence $E_k(\nu)$ is positive semidefinite.

Next, when $\lambda_{\min}(H_k) < 0$, it follows from (4.2.1) and Lemma 4.2.1 that

$$\lambda_{\min}(H_k) = \lambda_{\min}(P_k^T Q_k D_k Q_k^T P_k) \geq \lambda_{\min}(D_k)\lambda_{\max}(P_k^T Q_k Q_k^T P_k).$$

On the other hand, when $\lambda_{\min}(H_k) \geq 0$, we have $\lambda_{\min}(D_k) \geq 0$. Thus, we have

$$\max(0, -\lambda_{\min}(D_k)) \geq \frac{\Lambda_k}{\lambda_{\max}(P_k^T Q_k Q_k^T P_k)}.$$

It then follows from (4.2.2) and Lemma 4.2.1 that

$$\begin{aligned}
\lambda_{\min}(H_k + E_k(\nu)) &= \lambda_{\min}\left(P_k^T Q_k \Big(D_k + (c\max(0, -\lambda_{\min}(D_k)) + \nu\|g^k\|^\delta)I\Big) Q_k^T P_k\right) \\
&\geq \lambda_{\min}(P_k^T Q_k Q_k^T P_k)\lambda_{\min}(D_k + (c\max(0, -\lambda_{\min}(D_k)) + \nu\|g^k\|^\delta)I) \\
&\geq ((c-1)\max(0, -\lambda_{\min}(D_k)) + \nu\|g^k\|^\delta)\lambda_{\min}(P_k^T Q_k Q_k^T P_k) \\
&\geq \left(\frac{(c-1)\Lambda_k}{\lambda_{\max}(P_k^T Q_k Q_k^T P_k)} + \nu\|g^k\|^\delta\right)\lambda_{\min}(P_k^T Q_k Q_k^T P_k) \\
&\geq (3.781)^{2-2n}\left(\frac{(c-1)\Lambda_k}{4n^2 - 3n} + \nu\|g^k\|^\delta\right),
\end{aligned}$$

where the last inequality follows from Lemma 4.2.2 and the fact that P_k is a permutation matrix. Therefore, $E_k(\nu)$ holds Assumption 4.2.1 (b).

When $\lambda_{\min}(H_k) < 0$, it follows from (4.2.1) and Lemma 4.2.1 that

$$\lambda_{\min}(H_k) = \lambda_{\min}(P_k^T Q_k D_k Q_k^T P_k) \leq \lambda_{\min}(D_k)\lambda_{\min}(P_k^T Q_k Q_k^T P_k).$$

Thus, from (4.2.3), we have

$$\begin{aligned}
\|E_k(\nu)\| &\leq (c\max(0, -\lambda_{\min}(D)) + \nu\|g^k\|^\delta)\lambda_{\max}(P_k^T Q_k Q_k^T P_k) \\
&\leq \left(\frac{c\Lambda_k}{\lambda_{\min}(P_k^T Q_k Q_k^T P_k)} + \nu\|g^k\|^\delta\right)\lambda_{\max}(P_k^T Q_k Q_k^T P_k) \\
&\leq (4n^2 - 3n)\left(\frac{c\Lambda_k}{(3.781)^{2-2n}} + \nu\|g^k\|^\delta\right),
\end{aligned}$$

where the last inequality follows from Lemma 4.2.2 and the fact that P_k is a permutation matrix. Therefore, $E_k(\nu)$ holds Assumption 4.2.1 (c). □

In order to satisfy $f_{k+1} < f_k$ with $x^{k+1} := x^k + d^k(\nu_k)$, we propose to control ν_k. To find an appropriate ν_k, we use the idea of updating trust-region radius in the trust-region NM [17]. Let $\phi_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a model function of f at x^k defined by

$$\phi_k(d, \nu) := f_k + {g^k}^T d + \frac{1}{2} d^T (H_k + E_k(\nu))d.$$

Note that, $d^k(\nu)$ is a global minimizer of $\phi_k(\cdot, \nu)$ if $\|g^k\| \neq 0$. Let $\rho_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the ratio of the reduction of the objective function value to that of the model function value, i.e.,

$$\rho_k(d, \nu) := \frac{f_k - f(x^k + d)}{f_k - \phi_k(d, \nu)}.$$

If $\rho_k(d^k(\nu_k), \nu_k)$ is large, i.e., the reduction $f_k - f(x^k + d^k(\nu_k))$ is sufficiently large as compared to the reduction of the model function, then we adopt $d^k(\nu_k)$ and decrease the parameter ν_k. On the other hand, if $\rho_k(d^k(\nu_k), \nu_k)$ is small, i.e., the reduction $f_k - f(x^k + d^k(\nu_k))$ is not large, we increase ν_k and compute $d^k(\nu_k)$ once again.

The precise description of the proposed algorithm is as follows. We call the proposed algorithm the adaptive RNM since it uses an adaptive parameter ν.

The Adaptive Regularized Newton Method

Step 0 : Choose parameters $\nu_0, \nu_{\min}, \delta, \gamma_1, \gamma_2, \eta_1, \eta_2$ such that

$$0 < \nu_{\min} \leq \nu_0,\ \delta \geq 0,\ 0 < \gamma_1 \leq 1 < \gamma_2,\ 0 < \eta_1 < \eta_2 \leq 1.$$

Choose a starting point x^0. Set $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\nu}_{l_k} := \nu_k$.

Step 2.1 : Compute $d^k(\bar{\nu}_{l_k})$.

Step 2.2 : Compute $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})$. If $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1$, then set $\bar{\nu}_{l_k+1} := \gamma_2 \bar{\nu}_{l_k}$ and $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $\eta_1 \leq \rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_2$, then set $\nu_{k+1} := \bar{\nu}_{l_k}$. If $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_2$, then set $\nu_{k+1} := \max(\nu_{\min}, \gamma_1 \bar{\nu}_{l_k})$. Set $x^{k+1} := x^k + d^k(\bar{\nu}_{l_k})$ and $k := k + 1$. Go to Step 1.

In what follows, for simplicity, we denote l_k and $\bar{\nu}_{l_k}$ of the last iteration in the inner loops of Steps 2.0 – 2.2 at each k as l_k^* and ν_k^*, respectively.

In the remainder of this section, we show that the adaptive RNM is well-defined when $\|g^k\| \neq 0$. First, we give the following equality.

Lemma 4.2.3. *It satisfies that*

$$\begin{aligned} & f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu), \nu)) \\ & = \frac{2 - \eta_1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) - \frac{1}{2} \int_0^1 d^k(\nu)^T \nabla^2 f(x^k + \tau d^k(\nu)) d^k(\nu) \mathrm{d}\tau. \end{aligned}$$

Proof. By the definitions of $\phi_k(d^k(\nu), \nu)$ and $d^k(\nu)$, we have

$$\begin{aligned} f_k - \phi_k(d^k(\nu), \nu) &= -{g^k}^T d^k(\nu) - \frac{1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) \\ &= \frac{1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu). \end{aligned} \tag{4.2.4}$$

Since f is continuously differentiable, we have

$$\begin{aligned}
&f_k - f(x^k + d^k(\nu)) \\
&= -{g^k}^T d^k(\nu) - \frac{1}{2}\int_0^1 d^k(\nu)^T \nabla^2 f(x^k + \tau d^k(\nu)) d^k(\nu) \mathrm{d}\tau \\
&= d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) - \frac{1}{2}\int_0^1 d^k(\nu)^T \nabla^2 f(x^k + \tau d^k(\nu)) d^k(\nu) \mathrm{d}\tau,
\end{aligned}$$

where the second equality follows from the definition of $d^k(\nu)$. It then follows (4.2.4) that

$$\begin{aligned}
&f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu),\nu)) \\
&= \frac{2-\eta_1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) - \frac{1}{2}\int_0^1 d^k(\nu)^T \nabla^2 f(x^k + \tau d^k(\nu)) d^k(\nu) \mathrm{d}\tau,
\end{aligned}$$

which is the desired inequality. □

By using Lemma 4.2.3, we give the following inequality.

Lemma 4.2.4. *Suppose that Assumption* 4.2.1 *holds. Then,*

$$\begin{aligned}
&f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu),\nu)) \\
&\geq \frac{1}{2}\left((1-\eta_1)c_2\nu\|g^k\|^\delta - \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau\right)\|d^k(\nu)\|^2.
\end{aligned}$$

Proof. From Lemma 4.2.3 and the positive semidefiniteness of $E_k(\nu)$, we have

$$\begin{aligned}
&f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu),\nu)) \\
&\geq \frac{1-\eta_1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) - \frac{1}{2}\int_0^1 d^k(\nu)^T (\nabla^2 f(x^k + \tau d^k(\nu)) - H_k) d^k(\nu) \mathrm{d}\tau \\
&\geq \frac{1}{2}\left((1-\eta_1)\lambda_{\min}(H_k + E_k(\nu)) - \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau\right)\|d^k(\nu)\|^2 \\
&\geq \frac{1}{2}\left((1-\eta_1)(c_1\Lambda_k + c_2\nu\|g^k\|^\delta) - \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau\right)\|d^k(\nu)\|^2 \\
&\geq \frac{1}{2}\left((1-\eta_1)c_2\nu\|g^k\|^\delta - \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau\right)\|d^k(\nu)\|^2,
\end{aligned}$$

where the third inequality follows from Assumption 4.2.1 (b). □

Next, we give the following key lemma for the well-definedness.

Lemma 4.2.5. *Suppose that Assumption* 4.2.1 *holds. Suppose also that* $\|g^k\| \neq 0$. *Then,*

$$\rho_k(d^k(\nu),\nu) \geq \eta_1$$

for ν *sufficiently large.*

Proof. Taking $\nu \to \infty$, we have $\lim_{\nu\to\infty} \|d^k(\nu)\| = 0$ from the definition of $d^k(\nu)$ and Assumption 4.2.1, and hence

$$\lim_{\nu\to\infty} \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau = 0.$$

Thus, since $\|g^k\| \neq 0$, the following inequality holds for sufficiently large ν.

$$\int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \mathrm{d}\tau \leq (1-\eta_1)c_2\nu\|g^k\|^\delta.$$

It then follows from Lemma 4.2.4 that

$$f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu), \nu)) \geq 0$$

Therefore, by the definition of $\rho_k(d^k(\nu), \nu)$, it implies that $\rho_k(d^k(\nu), \nu) \geq \eta_1$. □

Now, we show that the proposed algorithm is well-defined.

Theorem 4.2.2. *Suppose that Assumption* 4.2.1 *holds. Suppose also that* $\|g^k\| \neq 0$. *Then the proposed algorithm is well-defined, i.e., the number* l_k *of inner iterations is finite.*

Proof. From the updating rule of $\bar{\nu}_{l_k}$, we have $\bar{\nu}_{l_k} \to \infty$ as $l_k \to \infty$. Thus, when l_k is sufficiently large, we have from Lemma 4.2.5 that $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$. Therefore, the proposed algorithm is well-defined. □

4.3 Global convergence

In this section, we investigate the global convergence property of the proposed algorithm. To this end, we need the following assumption.

Assumption 4.3.1.

(a) $\delta \leq 1$.

(b) *The level set of* f *at the initial point* x^0 *is compact, i.e.,* $\Omega := \{x \in \mathbb{R}^n \mid f(x) \leq f_0\}$ *is compact.*

Since $\{f_k\}$ is monotonically decreasing, the sequence $\{x^k\}$ is included in the compact set Ω under Assumption 4.3.1. Then, there exists $f_{\min}$ such that

$$f_k \geq f_{\min}, \quad \forall k \geq 0. \tag{4.3.1}$$

Moreover, since f is twice continuously differentiable, f is Lipschitz continuous on Ω. Thus, there exists a positive constant L_f such that

$$\|g^k\| \leq L_f, \quad \forall k \geq 0. \tag{4.3.2}$$

Now, we give an upper bound of $\|d^k(\nu)\|$.

Lemma 4.3.1. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Then, for any* $\nu \in [\nu_{\min}, \infty)$,

$$\|d^k(\nu)\| \le \frac{\|g^k\|^{1-\delta}}{c_2 \nu} \le U_d,$$

where

$$U_d := \frac{L_f^{1-\delta}}{c_2 \nu_{\min}}.$$

Proof. From the definition of $d^k(\nu)$, we have

$$\begin{aligned}
\|d^k(\nu)\| &= \|(H_k + E_k(\nu))^{-1} g^k\| \\
&\le \|(H_k + E_k(\nu))^{-1}\| \cdot \|g^k\| \\
&= \lambda_{\max}((H_k + E_k(\nu))^{-1}) \|g^k\| \\
&= \frac{\|g^k\|}{\lambda_{\min}(H_k + E_k(\nu))} \\
&\le \frac{\|g^k\|}{c_1 \Lambda_k + c_2 \nu \|g^k\|^\delta} \\
&\le \frac{\|g^k\|^{1-\delta}}{c_2 \nu},
\end{aligned} \tag{4.3.3}$$

where the second inequality follows from Assumption 4.2.1 (b). Then, from (4.3.2) and the fact that $\nu \ge \nu_{\min}$, we have

$$\|d^k(\nu)\| \le \frac{L_f^{1-\delta}}{c_2 \nu_{\min}}.$$

This completes the proof. □

From Lemma 4.3.1 we have

$$x^k + \tau d^k(\nu) \in \Omega + B(0, U_d), \quad \forall \tau \in [0,1], \quad \forall \nu \in [\nu_{\min}, \infty), \quad \forall k \ge 0. \tag{4.3.4}$$

Moreover, since $\Omega + B(0, U_d)$ is compact and f is twice continuously differentiable, ∇f is Lipschitz continuous on $\Omega + B(0, U_d)$. Thus, there exists a positive constant L_g such that

$$\|\nabla^2 f(x)\| \le L_g, \quad \forall x \in \Omega + B(0, U_d). \tag{4.3.5}$$

Next, we give the following inequality .

Lemma 4.3.2. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Then, for any* $\nu \in [\nu_{\min}, \infty)$,

$$f_k - f(x^k + d^k(\nu)) - \eta_1 (f_k - \phi_k(d^k(\nu), \nu)) \ge \frac{1}{2} \left((2 - \eta_1) c_2 \nu \|g^k\|^\delta - L_g \right) \|d^k(\nu)\|^2.$$

Proof. From Lemma 4.2.3, we have

$$
\begin{aligned}
& f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu), \nu)) \\
& \geq \frac{1}{2}\left((2-\eta_1)\lambda_{\min}(H_k + E_k(\nu)) - \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu))\| \mathrm{d}\tau\right) \|d^k(\nu)\|^2 \\
& \geq \frac{1}{2}\left((2-\eta_1)(c_1\Lambda_k + c_2\nu\|g^k\|^\delta) - L_g\right) \|d^k(\nu)\|^2 \\
& \geq \frac{1}{2}\left((2-\eta_1)c_2\nu\|g^k\|^\delta - L_g\right) \|d^k(\nu)\|^2.
\end{aligned}
$$

where the second inequality follows from (4.3.5) and Assumption 4.2.1 (b). □

By using Lemma 4.3.2, we show that $\rho_k(d^k(\nu), \nu) \geq \eta_1$ if ν is greater than a specific value depending on $\|g^k\|$.

Lemma 4.3.3. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Suppose also that*

$$
\nu \geq \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right),
$$

Then,

$$
\rho_k(d^k(\nu), \nu) \geq \eta_1.
$$

Proof. From Lemma 4.3.2 and the assumption on ν, we have

$$
f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu), \nu)) \geq 0.
$$

Therefore, by the definition of $\rho_k(d^k(\nu), \nu)$, it implies that $\rho_k(d^k(\nu), \nu) \geq \eta_1$. □

By using Lemma 4.3.3, we show that the parameter ν_k^* is bounded from above when $\|g^k\| \geq \varepsilon_g$ for all $k \geq 0$.

Lemma 4.3.4. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Suppose also that there exists a positive constant* ε_g *such that* $\|g^k\| \geq \varepsilon_g$ *for all* $k \geq 0$. *Then,*

$$
\nu_k^* \leq U_\nu(\varepsilon_g),
$$

where

$$
U_\nu(\varepsilon_g) := \gamma_2 \max\left(\frac{L_g}{(2-\eta_1)c_2\varepsilon_g^\delta}, \nu_0\right).
$$

Proof. From Lemma 4.3.3, if $\bar{\nu}_{l_k} \geq \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right)$, then $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$, and hence the inner loops of Step 2 must terminate. Therefore, if $\bar{\nu}_1 \geq \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right)$ at the k-th iteration, then $\nu_k^* = \bar{\nu}_1$. On the other hand, if $\bar{\nu}_1 < \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right)$, then ν_k^* must satisfy

$\nu_k^* \leq \gamma_2 \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right)$. Otherwise, $\bar{\nu}_{l_k^*-1} > \max\left(\frac{L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \nu_0\right)$, which contradicts $\rho_k(d^k(\bar{\nu}_{l_k^*-1}), \bar{\nu}_{l_k^*-1}) < \eta_1$. Consequently, we have

$$\begin{aligned}
\nu_k^* &\leq \max\left(\bar{\nu}_1, \frac{\gamma_2 L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \gamma_2\nu_0\right) \\
&= \max\left(\nu_{k-1}^*, \frac{\gamma_2 L_g}{(2-\eta_1)c_2\|g^k\|^\delta}, \gamma_2\nu_0\right) \\
&\leq \max\left(\nu_{k-1}^*, \frac{\gamma_2 L_g}{(2-\eta_1)c_2\varepsilon_g^\delta}, \gamma_2\nu_0\right) \\
&\leq \cdots \leq \max\left(\nu_0, \frac{\gamma_2 L_g}{(2-\eta_1)c_2\varepsilon_g^\delta}, \gamma_2\nu_0\right) \\
&= \gamma_2 \max\left(\frac{L_g}{(2-\eta_1)c_2\varepsilon_g^\delta}, \nu_0\right)
\end{aligned}$$

from the updating rule of ν. □

Next, we give a lower bound of the reduction of the model function.

Lemma 4.3.5. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Then, for any* $\nu \in [\nu_{\min}, \infty)$,

$$f_k - \phi_k(d^k(\nu), \nu) \geq \frac{1}{2\left((1+c_3)L_g + c_4\nu L_f^\delta\right)}\|g^k\|^2.$$

Proof. Since $H_k + E_k(\nu)$ is positive semidefinite, we have

$$\begin{aligned}
\lambda_{\min}((H_k + E_k(\nu))^{-1}) &= \frac{1}{\lambda_{\max}(H_k + E_k(\nu))} \geq \frac{1}{\|H_k\| + \|E_k(\nu)\|} \\
&\geq \frac{1}{\|H_k\| + c_3\Lambda_k + c_4\nu\|g^k\|^\delta} \geq \frac{1}{(1+c_3)L_g + c_4\nu L_f^\delta},
\end{aligned}$$

where the second inequality follows from Assumption 4.2.1 (c), and the last inequality follows from (4.3.2) and (4.3.5). It then follows from the definitions of $\phi_k(d^k(\nu), \nu)$ and $d^k(\nu)$ that

$$\begin{aligned}
f_k - \phi_k(d^k(\nu), \nu) &= -{g^k}^T d^k(\nu) - \frac{1}{2}d^k(\nu)^T(H_k + E_k(\nu))d^k(\nu) \\
&= \frac{1}{2}{g^k}^T(H_k + E_k(\nu))^{-1}g^k \\
&\geq \frac{1}{2}\lambda_{\min}((H_k + E_k(\nu))^{-1})\|g^k\|^2 \\
&\geq \frac{1}{2\left((1+c_3)L_g + c_4\nu L_f^\delta\right)}\|g^k\|^2,
\end{aligned}$$

which is the desired inequality. □

By using the above lemma, we give a lower bound of the reduction $f_k - f_{k+1}$ when $\|g^k\| \geq \varepsilon_g$ for all $k \geq 0$.

Lemma 4.3.6. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Suppose also that there exists a positive constant* ε_g *such that* $\|g^k\| \geq \varepsilon_g$ *for all* $k \geq 0$. *Then,*

$$f_k - f_{k+1} \geq p(\varepsilon_g)\varepsilon_g^2,$$

where

$$p(\varepsilon_g) := \frac{\eta_1}{2\left((1+c_3)L_g + c_4 U_\nu(\epsilon) L_f^\delta\right)}.$$

Proof. Since $\rho_k(d^k(\nu_k^*), \nu_k^*) \geq \eta_1$ from the definition of ν_k^*, we have

$$\begin{aligned} f_k - f_{k+1} &\geq \eta_1(f_k - \phi_k(d^k(\nu_k^*), \nu_k^*)) \\ &\geq \frac{\eta_1}{2\left((1+c_3)L_g + c_4\nu_k^* L_f^\delta\right)}\|g^k\|^2 \\ &\geq \frac{\eta_1}{2\left((1+c_3)L_g + c_4 U_\nu(\epsilon) L_f^\delta\right)}\varepsilon_g^2, \end{aligned}$$

where the second inequality follows from Lemma 4.3.5, and the last inequality follows from Lemma 4.3.4 and the fact that $\|g^k\| \geq \varepsilon_g$. □

Now, we are at the position to prove the main theorem of this section.

Theorem 4.3.1. *Suppose that Assumptions* 4.2.1 *and* 4.3.1 *hold. Then,*

$$\liminf_{k\to\infty} \|g^k\| = 0 \quad \text{or} \quad \|g^K\| = 0, \;\; \text{for some } K \geq 0.$$

Proof. Suppose the contrary, i.e., there exists a positive constant ε_g such that $\|g^k\| \geq \varepsilon_g$ for all $k \geq 0$. It then follows from Lemma 4.3.6 that

$$f_0 - f_k \geq \sum_{j=0}^{k-1}(f_j - f_{j+1}) \geq \sum_{j=0}^{k-1} p(\varepsilon_g)\varepsilon_g^2 = p(\varepsilon_g)\varepsilon_g^2 k.$$

Taking $k \to \infty$, the right-hand side of the inequality goes to infinity, and hence $\lim_{k\to\infty} f_k = -\infty$. This contradicts (4.3.1). Therefore, we have $\liminf_{k\to\infty}\|g^k\| = 0$ or $\|g^K\| = 0$ for some $K \geq 0$. □

Remark 4.3.1. *From the updating rule of* ν, *Lemma* 4.3.4 *requires the assumption that* $\|g^k\| \geq \varepsilon_g$ *for all* $k \geq 0$. *Thus, we cannot apply the proof of Theorem* 3.3.1, *where the global convergence property of the extended RNM is shown.*

4.4 Global complexity bound

In this section, we estimate the global complexity bounds of the adaptive RNM. Let K_{outer} be the total number of outer iterations until $\|g^k\| \leq \varepsilon_g$ holds for the first time. Let also $\widehat{K}_{\text{outer}}$ be the total number of outer iterations until $f_k - \inf_{x\in\mathbb{R}^n} f(x) \leq \varepsilon_f$ holds for the first time. If there do not exist such K_{outer} or $\widehat{K}_{\text{outer}}$, we define $K_{\text{outer}} := \infty$ or $\widehat{K}_{\text{outer}} := \infty$. Moreover, let K_{total} and $\widehat{K}_{\text{total}}$ be the total number of inner iterations, i.e.,

$$K_{\text{total}} := \sum_{k=0}^{K_{\text{outer}}-1} l_k^*,$$

$$\widehat{K}_{\text{total}} := \sum_{k=0}^{\widehat{K}_{\text{outer}}-1} l_k^*.$$

To investigate K_{total} and $\widehat{K}_{\text{total}}$, we need the following assumption throughout this section.

Assumption 4.4.1. *$\nabla^2 f$ is Lipschitz continuous on $\Omega + B(0, U_d)$ with modulus L_H.*

From (4.3.4) and Assumption 4.4.2, we have

$$\|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \leq L_H \tau \|d^k(\nu)\|, \quad \forall \tau \in [0,1], \quad \forall \nu \in [\nu_{\min}, \infty), \quad \forall k \geq 0. \tag{4.4.1}$$

In what follows, we consider three cases (a) f is nonconvex, (b) f is convex and (c) f is strongly convex.

4.4.1 Nonconvex case

In the case where f is nonconvex, we need the following assumptions in addition to Assumption 4.4.1.

Assumption 4.4.2. $\delta \leq 1/2$.

First, we show that $\rho_k(d^k(\nu), \nu) \geq \eta_1$ if ν is greater than a specific value independent of k.

Lemma 4.4.1. *Suppose that Assumptions 4.2.1, 4.3.1, 4.4.1 and 4.4.2 hold. Suppose also that*

$$\nu \geq \max\left(\left(\frac{L_H L_f^{1-2\delta}}{2c_2^2(1-\eta_1)} \right)^{\frac{1}{2}}, \nu_0 \right)$$

Then,

$$\rho_k(d^k(\nu), \nu) \geq \eta_1.$$

Proof. From Lemma 4.2.4 and (4.4.1), we have

$$\begin{aligned}
& f_k - f(x^k + d^k(\nu)) - \eta_1 (f_k - \phi_k(d^k(\nu), \nu)) \\
& \geq \frac{1}{2} \left((1-\eta_1) c_2 \nu \|g^k\|^\delta - L_H \|d^k(\nu)\| \int_0^1 \tau \mathrm{d}\tau \right) \|d^k(\nu)\|^2 \\
& = \frac{1}{2} \left((1-\eta_1) c_2 \nu \|g^k\|^\delta - \frac{L_H}{2} \|d^k(\nu)\| \right) \|d^k(\nu)\|^2 \\
& \geq \frac{1}{2\nu} \left((1-\eta_1) c_2 \nu^2 - \frac{L_H \|g^k\|^{1-2\delta}}{2c_2} \right) \|g^k\|^\delta \|d^k(\nu)\|^2 \\
& \geq \frac{1}{2\nu} \left((1-\eta_1) c_2 \nu^2 - \frac{L_H L_f^{1-2\delta}}{2c_2} \right) \|g^k\|^\delta \|d^k(\nu)\|^2 \\
& \geq 0,
\end{aligned} \tag{4.4.2}$$

where the second inequality follows from Lemma 4.3.1, the third inequality follows from (4.3.2), and the last inequality follows from the assumption on ν. Therefore, by the definition of $\rho_k(d^k(\nu),\nu)$, it implies that $\rho_k(d^k(\nu),\nu) \geq \eta_1$. □

The next lemma indicates that the parameter ν_k^* is bounded from above by some positive constant independent of k.

Lemma 4.4.2. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.2 *hold. Then,*

$$\nu_k^* \leq U_\nu,$$

where

$$U_\nu := \gamma_2 \max\left(\left(\frac{L_H L_f^{1-2\delta}}{2c_2^2(1-\eta_1)}\right)^{\frac{1}{2}}, \nu_0\right).$$

Proof. By using Lemma 4.4.1 and a way similar to the proof of Lemma 4.3.4, we obtain the desired inequality. □

From the above lemma, we show that the number l_k^* of inner iterations at the k-th iteration is bounded from above by some positive constant independent of k.

Theorem 4.4.1. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.2 *hold. Then,*

$$l_k^* \leq \left\lceil \log_{\gamma_2}\left(\frac{U_\nu}{\nu_{\min}}\right) + 1\right\rceil.$$

Proof. From Lemma 4.4.2, we have $\nu_{\min} \leq \bar{\nu}_{l_k} \leq U_\nu$. Therefore, by the updating rule of ν, we obtain the desired inequality. □

Next, we give a lower bound of the reduction $f_k - f_{k+1}$.

Lemma 4.4.3. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.2 *hold. Then,*

$$f_k - f_{k+1} \geq p_1 \|g^k\|^2,$$

where

$$p_1 := \frac{\eta_1}{2\left((1+c_3)L_g + c_4 U_\nu L_f^\delta\right)}.$$

Proof. By using Lemma 4.4.2 and a way similar to the proof of Lemmas 4.3.5 and 4.3.6, we obtain the desired inequality. □

By using Lemma 4.4.3, we give an upper bound of K_{outer}.

Theorem 4.4.2. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.2 *hold. Then,*

$$K_{\text{outer}} \leq \left\lceil \frac{f_0 - f_{\min}}{p_1}\varepsilon_g^{-2} + 1\right\rceil.$$

Proof. Let K be $\lceil((f_0 - f_{\min})\varepsilon_g^{-2}/p_1) + 1\rceil$. Suppose the contrary, i.e., $K_{\text{outer}} > K$. It then follows from (4.3.1) and Lemma 4.4.3 that

$$f_0 - f_{\min} \geq f_0 - f_K = \sum_{j=0}^{K-1}(f_j - f_{j+1}) > \sum_{j=0}^{K-1} p_1\varepsilon_g^2 = p_1\varepsilon_g^2 K. \tag{4.4.3}$$

On the other hand, we have

$$p_1\varepsilon_g^2 K = p_1\varepsilon_g^2 \left\lceil \left(\frac{f_0 - f_{\min}}{p_1\varepsilon^2}\right) + 1 \right\rceil > f_0 - f_{\min}$$

from the definition of K. This contradicts (4.4.3), and hence we obtain the theorem. □

Now, we give the global complexity bound K_{total}.

Theorem 4.4.3. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.2 *hold. Then,*

$$K_{\text{total}} \leq \left\lceil \log_{\gamma_2}\left(\frac{U_\nu}{\nu_{\min}}\right) + 1 \right\rceil \cdot \left\lceil \frac{f_0 - f_{\min}}{p_1}\varepsilon_g^{-2} + 1 \right\rceil.$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$.

Proof. From the definition of K_{total}, we have

$$\begin{aligned} K_{\text{total}} &= \sum_{k=0}^{K_{\text{outer}}-1} l_k^* \leq \sum_{k=0}^{K_{\text{outer}}-1} \left\lceil \log_{\gamma_2}\left(\frac{U_\nu}{\nu_{\min}}\right) + 1 \right\rceil \\ &= K_{\text{outer}} \left\lceil \log_{\gamma_2}\left(\frac{U_\nu}{\nu_{\min}}\right) + 1 \right\rceil \leq \left\lceil \log_{\gamma_2}\left(\frac{U_\nu}{\nu_{\min}}\right) + 1 \right\rceil \cdot \left\lceil \frac{f_0 - f_{\min}}{p_1}\varepsilon_g^{-2} + 1 \right\rceil, \end{aligned}$$

where the first inequality follows from Theorem 4.4.1, and the last inequality follows from Theorem 4.4.2. □

The above global complexity bound is the same as that of the steepest descent method. On the other hand, it can be reduced under the following additional assumption on the minimum eigenvalue of H_k.

Assumption 4.4.3. *There exist positive constants* $\bar{\delta}$ *and* c_5 *such that*

$$\Lambda_k \leq c_5\|g^k\|^{\bar{\delta}}, \quad \forall k \geq 0.$$

In what follows, let $\hat{\delta} := \min(\delta, \bar{\delta})$.

Before we show the reduced complexity bound, we give sufficient conditions for Assumption 4.4.3.

Proposition 4.4.1.

(a) *Suppose that f is convex. Then, Assumption* 4.4.3 *holds for any* $\bar{\delta}$ *and* c_4.

(b) *Suppose that Assumptions* 4.3.1 *and* 4.4.1 *hold. Suppose also that f is analytic and* $\nabla^2 f(x) \succeq 0$ *for any x such that* $\nabla f(x) = 0$. *Then, Assumption* 4.4.3 *holds.*

Proof. The statement (a) directly follows from the fact that $\Lambda_k = 0, \forall k \geq 0$ when f is convex.

Next, we show (b). Let $X_1 := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\| = 0\}$ and $X_2 := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\| = 0,\ \nabla^2 f(x) \succeq 0\}$. In a way similar to the proof of Lemma 3.5.2, we can show that

$$\Lambda_k \leq L_H \mathrm{dist}(x^k, X_2),$$

when Assumption 4.4.1 holds. Moreover, it is shown in [56] that there exist positive constants κ and δ' such that

$$\mathrm{dist}(x, X_1) \leq \kappa \|\nabla f(x)\|^{\delta'}, \quad \forall x \in \Omega,$$

when f is analytic and Ω is compact. It then follows from $X_1 = X_2$ that

$$\Lambda_k \leq \kappa L_H \|g^k\|^{\delta'},$$

and hence Assumption 4.4.3 holds. □

Remark 4.4.1. *If f is quasi-convex, then $\nabla^2 f(x) \succeq 0$ for any x such that $\nabla f(x) = 0$ [19].*

Now, we give a lower bound of $\|d^k(\nu)\|$.

Lemma 4.4.4. *Suppose that Assumptions 4.2.1 and 4.3.1 hold. Then, for any $\nu \in [\nu_{\min}, \infty)$,*

$$\|d^k(\nu)\| \geq \frac{1}{(1 + c_3)L_g + c_4 \nu L_f^{\delta}} \|g^k\|.$$

Proof. From the definition of $d^k(\nu)$, we have

$$\begin{aligned} \|g^k\| &= \|(H_k + E_k(\nu))d^k(\nu)\| \leq (\|H_k\| + \|E_k(\nu)\|)\|d^k(\nu)\| \\ &\leq (\|H_k\| + c_3\Lambda_k + c_4\nu\|g^k\|^{\delta})\|d^k(\nu)\| \leq ((1 + c_3)L_g + c_4\nu L_f^{\delta})\|d^k(\nu)\|, \end{aligned}$$

where the second inequality follows from Assumption 4.2.1 (c), and the last inequality follows from (4.3.2) and (4.3.5). □

Next, we give a lower bound of the reduction of the model function.

Lemma 4.4.5. *Suppose that Assumptions 4.2.1 hold. Then,*

$$f_k - \phi_k(d^k(\nu), \nu) \geq \frac{1}{2}(c_1\Lambda_k + c_2\nu\|g^k\|^{\delta})\|d^k(\nu)\|^2.$$

Proof. From the definitions of $\phi_k(d^k(\nu), \nu)$ and $d^k(\nu)$, we have

$$\begin{aligned} f_k - \phi_k(d^k(\nu), \nu) &= -{g^k}^T d^k(\nu) - \frac{1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) \\ &= \frac{1}{2} d^k(\nu)^T (H_k + E_k(\nu)) d^k(\nu) \\ &\geq \frac{1}{2}\lambda_{\min}(H_k + E_k(\nu))\|d^k(\nu)\|^2 \\ &\geq \frac{1}{2}(c_1\Lambda_k + c_2\nu\|g^k\|^{\delta})\|d^k(\nu)\|^2, \end{aligned}$$

where the last inequality follows from Assumption 4.2.1 (b). □

Next, we show the following key lemma.

Lemma 4.4.6. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.3 *hold. Then,*

$$\|g^{k+1}\| \le c_6 \max(\Lambda_k \|d^k(\nu_k^*)\|, \|g^k\|^\delta \|d^k(\nu_k^*)\|),$$

where

$$c_6 := c_3 + c_4 + \frac{L_H L_f^{1-2\delta}}{2c_1 \nu_{\min}}.$$

Proof. Since f is twice continuously differentiable, we have

$$g^{k+1} = g^k + \int_0^1 \nabla^2 f(x^k + \tau d^k(\nu_k^*)) d^k(\nu_k^*) \mathrm{d}\tau.$$

It then follows that

$$\begin{aligned}
\|g^{k+1} - g^k - H_k d^k(\nu_k^*)\| &= \left\| \int_0^1 (\nabla^2 f(x^k + \tau d^k(\nu_k^*)) - H_k) d^k(\nu_k^*) \mathrm{d}\tau \right\| \\
&\le \|d^k(\nu_k^*)\| \int_0^1 \|\nabla^2 f(x^k + \tau d^k(\nu_k^*)) - H_k\| \mathrm{d}\tau \\
&\le L_H \|d^k(\nu_k^*)\|^2 \int_0^1 \tau \mathrm{d}\tau \\
&= \frac{L_H}{2} \|d^k(\nu_k^*)\|^2,
\end{aligned}$$

where the second inequality follows from (4.4.1). Thus, we have

$$\begin{aligned}
\|g^{k+1}\| &\le \| - H_k d^k(\nu_k^*) - g^k\| + \frac{L_H}{2} \|d^k(\nu_k^*)\|^2 \\
&= \|E_k(\nu_k^*) d^k(\nu_k^*)\| + \frac{L_H}{2} \|d^k(\nu_k^*)\|^2 \\
&\le c_3 \Lambda_k \|d^k(\nu_k^*)\| + c_4 \|g^k\|^\delta \|d^k(\nu_k^*)\| + \frac{L_H}{2} \|d^k(\nu_k^*)\|^2 \\
&\le c_3 \Lambda_k \|d^k(\nu_k^*)\| + \left(c_4 + \frac{L_H}{2c_1 \nu_{\min}} \|g^k\|^{1-2\delta} \right) \|g^k\|^\delta \|d^k(\nu_k^*)\| \\
&\le c_3 \Lambda_k \|d^k(\nu_k^*)\| + \left(c_4 + \frac{L_H L_f^{1-2\delta}}{2c_1 \nu_{\min}} \right) \|g^k\|^\delta \|d^k(\nu_k^*)\| \\
&\le \left(c_3 + c_4 + \frac{L_H L_f^{1-2\delta}}{2c_1 \nu_{\min}} \right) \max(\Lambda_k \|d^k(\nu_k^*)\|, \|g^k\|^\delta \|d^k(\nu_k^*)\|),
\end{aligned} \tag{4.4.4}$$

where the first equality follows from the definition of $d^k(\nu_k^*)$, the second inequality follows from Assumption 4.2.1 (c), the third inequality follows from Assumption 4.4.3, Lemma 4.3.1, and the fact that $\nu_k^* \ge \nu_{\min}$, and the fourth inequality follows from (4.3.2). □

Now, we give a lower bound of the reduction $f_k - f_{k+1}$.

Lemma 4.4.7. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.3 *hold. Then,*

$$f_k - f_{k+1} \geq p_2 \|g^{k+1}\|^{\frac{2+\hat{\delta}}{1+\hat{\delta}}},$$

where

$$p_2 := \min\left(\frac{c_1\eta_1}{2c_5c_6^2 L_f^{\bar{\delta}-\hat{\delta}}}, \frac{c_2\eta_1\nu_{\min}}{2c_6^2 L_f^{\delta-\hat{\delta}}}, \frac{\eta_1 \min(c_1, c_2\nu_{\min})}{2c_6((1+c_3)L_g + c_4U_\nu L_f^{\delta})} \right).$$

Proof. Since $\rho_k(d^k(\nu_k^*), \nu_k^*) \geq \eta_1$ from the definition of ν_k^*, we have

$$\begin{aligned} f_k - f_{k+1} &\geq \eta_1(f_k - \phi_k(d^k(\nu_k^*), \nu_k^*)) \\ &\geq \frac{1}{2}\eta_1(c_1\Lambda_k + c_2\nu_k^*\|g^k\|^{\delta})\|d^k(\nu_k^*)\|^2 \\ &\geq \frac{1}{2}c_1\eta_1\Lambda_k\|d^k(\nu_k^*)\|^2, \end{aligned} \tag{4.4.5}$$

where the second inequality follows from Lemma 4.4.5. From Lemma 4.4.5, we also have

$$f_k - f_{k+1} \geq \frac{1}{2}c_2\eta_1\nu_k^*\|g^k\|^{\delta}\|d^k(\nu_k^*)\|^2 \geq \frac{1}{2}c_2\eta_1\nu_{\min}\|g^k\|^{\delta}\|d^k(\nu_k^*)\|^2, \tag{4.4.6}$$

where the last inequality follows from the fact that $\nu_k^* \geq \nu_{\min}$. In what follows, we consider two cases: (i) $\Lambda_k\|d^k(\nu_k^*)\| \geq \|g^k\|^{\delta}\|d^k(\nu_k^*)\|$ and (ii) $\Lambda_k\|d^k(\nu_k^*)\| \leq \|g^k\|^{\delta}\|d^k(\nu_k^*)\|$.

Case (i): In this case, from Lemma 4.4.6, we have

$$\|g^{k+1}\| \leq c_6\Lambda_k\|d^k(\nu_k^*)\|. \tag{4.4.7}$$

It then follows from (4.4.5) that

$$\begin{aligned} f_k - f_{k+1} &\geq \frac{c_1\eta_1}{2c_6^2\Lambda_k}\|g^{k+1}\|^2 \geq \frac{c_1\eta_1}{2c_5c_6^2\|g^k\|^{\bar{\delta}}}\|g^{k+1}\|^2 \\ &= \frac{c_1\eta_1}{2c_5c_6^2\|g^k\|^{\bar{\delta}-\hat{\delta}}}\|g^k\|^{-\hat{\delta}}\|g^{k+1}\|^2 \geq \frac{c_1\eta_1}{2c_5c_6^2 L_f^{\bar{\delta}-\hat{\delta}}}\|g^k\|^{-\hat{\delta}}\|g^{k+1}\|^2, \end{aligned} \tag{4.4.8}$$

where the second inequality follows from Assumption 4.4.3, and the last inequality follows from (4.3.2). On the other hand, from (4.4.5) and (4.4.7), we have

$$\begin{aligned} f_k - f_{k+1} &\geq \frac{c_1\eta_1}{2c_6}\|d^k(\nu_k^*)\| \cdot \|g^{k+1}\|^2 \\ &\geq \frac{c_1\eta_1}{2c_6((1+c_3)L_g + c_4U_\nu L_f^{\delta})}\|g^k\| \cdot \|g^{k+1}\|^2, \end{aligned} \tag{4.4.9}$$

where the second inequality follows from Lemma 4.4.4. Now, we consider two cases: (a) $\|g^{k+1}\| \geq \|g^k\|^{\alpha}$ and (b) $\|g^{k+1}\| \leq \|g^k\|^{\alpha}$, where α is an arbitrary positive constant.

Case (a): This case implies that

$$\|g^k\|^{-\hat{\delta}} \geq \|g^{k+1}\|^{-\frac{\hat{\delta}}{\alpha}}.$$

It then follows from (4.4.8) that

$$f_k - f_{k+1} \geq \frac{c_1\eta_1}{2c_5c_6^2 L_f^{\bar{\delta}-\hat{\delta}}}\|g^{k+1}\|^{2-\frac{\hat{\delta}}{\alpha}}, \tag{4.4.10}$$

Case (b): In this case, we have

$$\|g^k\| \geq \|g^{k+1}\|^{\frac{1}{\alpha}}.$$

It then follows from (4.4.9) that

$$f_k - f_{k+1} \geq \frac{c_1\eta_1}{2c_6((1+c_3)L_g + c_4U_\nu L_f^\delta)}\|g^{k+1}\|^{1+\frac{1}{\alpha}}. \tag{4.4.11}$$

Since α is an arbitrary positive constant, we choose $\alpha := 1+\hat{\delta}$, which minimizes $\max(2-\frac{\hat{\delta}}{\alpha}, 1+\frac{1}{\alpha})$. Then, we have

$$2 - \frac{\hat{\delta}}{\alpha} = 1 + \frac{1}{\alpha} = \frac{2+\hat{\delta}}{1+\hat{\delta}}.$$

It then follows from (4.4.10) and (4.4.11) that

$$f_k - f_{k+1} \geq \min\left(\frac{c_1\eta_1}{2c_5c_6^2 L_f^{\bar{\delta}-\hat{\delta}}}, \frac{c_1\eta_1}{2c_6((1+c_3)L_g + c_4U_\nu L_f^\delta)}\right)\|g^{k+1}\|^{\frac{2+\hat{\delta}}{1+\hat{\delta}}} \tag{4.4.12}$$

Case (ii): In this case, from Lemma 4.4.6, we have

$$\|g^{k+1}\| \leq c_6\|g^k\|^\delta\|d^k(\nu_k^*)\|. \tag{4.4.13}$$

It then follows from (4.4.6) that

$$\begin{aligned} f_k - f_{k+1} &\geq \frac{c_2\eta_1\nu_{\min}}{2c_6^2\|g^k\|^\delta}\|g^{k+1}\|^2 = \frac{c_2\eta_1\nu_{\min}}{2c_6^2\|g^k\|^{\delta-\hat{\delta}}}\|g^k\|^{-\hat{\delta}}\|g^{k+1}\|^2 \\ &\geq \frac{c_2\eta_1\nu_{\min}}{2c_6^2 L_f^{\delta-\hat{\delta}}}\|g^k\|^{-\hat{\delta}}\|g^{k+1}\|^2, \end{aligned} \tag{4.4.14}$$

where the last inequality follows from (4.3.2). On the other hand, from (4.4.6) and (4.4.13), we have

$$\begin{aligned} f_k - f_{k+1} &\geq \frac{c_2\eta_1\nu_{\min}}{2c_6}\|d^k(\nu_k^*)\| \cdot \|g^{k+1}\|^2 \\ &\geq \frac{c_2\eta_1\nu_{\min}}{2c_6((1+c_3)L_g + c_4U_\nu L_f^\delta)}\|g^k\| \cdot \|g^{k+1}\|^2, \end{aligned} \tag{4.4.15}$$

where the second inequality follows from Lemma 4.4.4. Therefore, by using (4.4.14), (4.4.15) and the technique in the case (i), we have

$$f_k - f_{k+1} \geq \min\left(\frac{c_2\eta_1\nu_{\min}}{2c_6^2 L_f^{\delta-\hat{\delta}}}, \frac{c_2\eta_1\nu_{\min}}{2c_6((1+c_3)L_g + c_4U_\nu L_f^\delta)}\right)\|g^{k+1}\|^{\frac{2+\hat{\delta}}{1+\hat{\delta}}} \tag{4.4.16}$$

Finally, from (4.4.12) and (4.4.16), we have

$$f_k - f_{k+1} \geq \min\left(\frac{c_1\eta_1}{2c_5c_6^2L_f^{\bar{\delta}-\hat{\delta}}}, \frac{c_2\eta_1\nu_{\min}}{2c_6^2L_f^{\delta-\hat{\delta}}}, \frac{\eta_1\min(c_1, c_2\nu_{\min})}{2c_6((1+c_3)L_g + c_4U_\nu L_f^{\delta})}\right)\|g^{k+1}\|^{\frac{2+\hat{\delta}}{1+\hat{\delta}}},$$

which is the desired inequality. □

From Lemma 4.4.7, we give an upper bound of K_{outer}.

Theorem 4.4.4. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.3 *hold. Then,*

$$K_{\text{outer}} \leq \left\lceil \frac{f_0 - f_{\min}}{p_2}\varepsilon_g^{-\frac{2+\hat{\delta}}{1+\hat{\delta}}} + 1 \right\rceil.$$

Proof. By using Lemma 4.4.7 and a way similar to the proof of Theorem 4.4.2, we obtain the desired inequality. □

From Theorems 4.4.1 and 4.4.4, we can directly give the following global complexity bound K_{total}.

Theorem 4.4.5. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.3 *hold. Then,*

$$K_{\text{total}} = O(\varepsilon_g^{-\frac{2+\hat{\delta}}{1+\hat{\delta}}}).$$

Remark 4.4.2. *Under Assumption* 4.4.3*, the global complexity bound* $O(\varepsilon_g^{-\frac{2+\hat{\delta}}{1+\hat{\delta}}})$ *of the proposed algorithm is better than* $O(\varepsilon_g^{-2})$ *of the steepest descent method.*

4.4.2 Convex case

In this subsection, we consider the case where f is convex.

Assumption 4.4.4. f *is convex.*

From Proposition 4.4.1 (a), Assumption 4.4.3 holds for any c_5 and $\bar{\delta}$. Then, we have $\hat{\delta} = \delta$. Moreover, under Assumptions 4.4.2 and 4.4.4, Theorems 4.4.1 and 4.4.5 hold. Thus, we can directly give the following global complexity bound K_{total}.

Theorem 4.4.6. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.4 *hold. Then,*

$$K_{\text{total}} = O(\varepsilon_g^{-\frac{2+\delta}{1+\delta}}).$$

In particular, if $\delta = 1/2$*, then*

$$K_{\text{total}} = O(\varepsilon_g^{-\frac{5}{3}}).$$

In what follows, we discuss the global complexity bound $\widehat{K}_{\text{total}}$. From Assumption 4.3.1, there exists a solution x^* of (4.1.1). Moreover, there exists a positive constant U_x such that

$$\|x^k - x^*\| \le U_x, \quad \forall k \ge 0. \tag{4.4.17}$$

First, we give the following technical lemma.

Lemma 4.4.8. *Let β, γ and u be positive parameters such that $0 < \beta \le 1$, $\gamma \ge 0$ and $u > 0$. Then,*

$$(1+\gamma\alpha)^\beta \ge 1 + \frac{(1+\gamma u)^\beta - 1}{u}\alpha, \quad \forall \alpha \in [0, u]. \tag{4.4.18}$$

Proof. Let $h(t) := (1+\gamma t)^\beta$. Since $0 < \beta \le 1$ and $\gamma \ge 0$, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} h(t) = -\frac{\beta(1-\beta)\gamma^2}{(1+\gamma t)^{2-\beta}} \le 0, \quad \forall t \in [0, \infty)$$

Therefore, $h(t)$ is concave on $[0, u]$. Let $\alpha \in [0, u]$. Then, $\alpha/u \in [0, 1]$. It then follows from the concavity of h that

$$h(\alpha) = h\left(\frac{\alpha}{u}u + \left(1 - \frac{\alpha}{u}\right)0\right) \ge \frac{\alpha}{u}h(u) + \left(1 - \frac{\alpha}{u}\right)h(0) = 1 + \frac{(1+\gamma u)^\beta - 1}{u}\alpha,$$

which is the desired inequality. □

By using Lemma 4.4.8, we give an upper bound of $\widehat{K}_{\text{outer}}$. The proof technique is based on the technique in [53], where an upper bound of $\widehat{K}_{\text{outer}}$ for the cubic regularization of the Newton method is given.

Theorem 4.4.7. *Suppose that Assumptions* 4.2.1*,* 4.3.1*,* 4.4.1*,* 4.4.2 *and* 4.4.4 *hold. Then,*

$$\widehat{K}_{\text{outer}} = O(\varepsilon_f^{-\frac{1}{1+\delta}}).$$

Proof. From Assumption 4.4.4 and (4.4.17), we have

$$f_{k+1} - f(x^*) \le {g^{k+1}}^T(x^{k+1} - x^*) \le U_x \|g^{k+1}\|.$$

It then follows from Lemma 4.4.7 that

$$f_k - f_{k+1} \ge \frac{p_2}{U_x^{\frac{2+\delta}{1+\delta}}} \left(f_{k+1} - f(x^*)\right)^{\frac{2+\delta}{1+\delta}}.$$

Denoting $\alpha_k := f_k - f(x^*)$, $\beta := 1/(1+\delta)$ and $\gamma := \eta_1 p_2 / U_x^{\frac{2+\delta}{1+\delta}}$, we obtain

$$\alpha_k \ge \alpha_{k+1} + \gamma \alpha_{k+1}^{1+\beta}.$$

Then, we have

$$\begin{aligned}\frac{1}{\alpha_{k+1}^{\beta}} - \frac{1}{\alpha_k^{\beta}} &\geq \frac{1}{\alpha_{k+1}^{\beta}} - \frac{1}{(\alpha_{k+1} + \gamma\alpha_{k+1}^{1+\beta})^{\beta}} \\ &= \frac{\alpha_{k+1}^{\beta}(1+\gamma\alpha_{k+1}^{\beta})^{\beta} - \alpha_{k+1}^{\beta}}{\alpha_{k+1}^{2\beta}(1+\gamma\alpha_{k+1}^{\beta})^{\beta}} \\ &= \frac{(1+\gamma\alpha_{k+1}^{\beta})^{\beta} - 1}{\alpha_{k+1}^{\beta}(1+\gamma\alpha_{k+1}^{\beta})^{\beta}}. \end{aligned} \tag{4.4.19}$$

Since $\alpha_{k+1}^{\beta} \leq \alpha_0^{\beta}$ and $\beta \leq 1$, substituting $u := \alpha_0^{\beta}$ and $\alpha := \alpha_{k+1}^{\beta}$ into (4.4.18) of Lemma 4.4.8 yields

$$1 + \frac{(1+\gamma\alpha_0^{\beta})^{\beta} - 1}{\alpha_0^{\beta}}\alpha_{k+1}^{\beta} \leq (1+\gamma\alpha_{k+1}^{\beta})^{\beta} \leq (1+\gamma\alpha_0^{\beta})^{\beta}.$$

It then follows from (4.4.7) that

$$\begin{aligned}\frac{1}{\alpha_{k+1}^{\beta}} &\geq \frac{1}{\alpha_k^{\beta}} + \frac{(1+\gamma\alpha_0^{\beta})^{\beta} - 1}{\alpha_0^{\beta}(1+\gamma\alpha_0^{\beta})^{\beta}} \\ &\geq \frac{1}{\alpha_0^{\beta}} + \frac{(1+\gamma\alpha_0^{\beta})^{\beta} - 1}{\alpha_0^{\beta}(1+\gamma\alpha_0^{\beta})^{\beta}}(k+1) \\ &= \frac{(1+\gamma\alpha_0^{\beta})^{\beta} + \left((1+\gamma\alpha_0^{\beta})^{\beta} - 1\right)(k+1)}{\alpha_0^{\beta}(1+\gamma\alpha_0^{\beta})^{\beta}}, \end{aligned}$$

and hence

$$\alpha_k \leq \left(\frac{\alpha_0^{\beta}(1+\gamma\alpha_0^{\beta})^{\beta}}{(1+\gamma\alpha_0^{\beta})^{\beta} + \left((1+\gamma\alpha_0^{\beta})^{\beta} - 1\right)k}\right)^{\frac{1}{\beta}}.$$

Therefore, $f_k - f(x^*) = \alpha_k \leq \varepsilon_f$, provided that

$$k \geq \frac{\alpha_0^{\beta}(1+\gamma\alpha_0^{\beta})^{\beta}\varepsilon_f^{-\beta} - (1+\gamma\alpha_0^{\beta})^{\beta}}{(1+\gamma\alpha_0^{\beta})^{\beta} - 1}.$$

This completes the proof. □

From Theorems 4.4.1 and 4.4.7, we can directly give the following global complexity bound $\widehat{K}_{\text{total}}$.

Theorem 4.4.8. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1, 4.4.2 *and* 4.4.4 *hold. Then,*

$$\widehat{K}_{\text{total}} = O(\varepsilon_f^{-\frac{1}{1+\delta}}).$$

In particular, if $\delta = 1/2$, *then*

$$\widehat{K}_{\text{total}} = O(\varepsilon_f^{-\frac{2}{3}}).$$

Remark 4.4.3. *The global complexity bounds* $K_{\text{total}} = O(\varepsilon_g^{-\frac{2+\delta}{1+\delta}})$ *and* $\widehat{K}_{\text{total}} = O(\varepsilon_f^{-\frac{1}{1+\delta}})$ *become better as we take a larger* δ. *However, we need* $\delta \leq 1/2$ *for Lemma* 4.4.1. *Thus, the upper bounds of* K_{total} *and* $\widehat{K}_{\text{total}}$ *are* $O(\varepsilon_g^{-\frac{5}{3}})$ *and* $O(\varepsilon_f^{-\frac{2}{3}})$, *respectively.*

4.4.3 Strongly convex case

In this subsection, we give the global complexity bound in the case where f is strongly convex. Moreover, we show that a sequence $\{f_k - f(x^*)\}$ globally linearly converges to 0 as well as the steepest descent method [50] and the cubic regularization of the Newton method [53].

From Remark 4.4.3, we expect to get the better global complexity bound as we take a larger δ. Therefore, it is worth considering the case where $\delta > 1/2$. When $\delta > 1/2$, Lemma 4.4.1 does not always hold. However, when f is strongly convex, we can relax the assumption $\delta \leq 1/2$ to $\delta \leq 1$, and prove properties similar to Lemma 4.4.1.

Now, we formally state assumptions used in this subsection.

Assumption 4.4.5.

(a) $\delta \leq 1$.

(b) *f is strongly convex, i.e., there exists a positive constant σ such that*

$$f(x) \geq f(y) + \nabla f(y)^T(x-y) + \frac{1}{2}\sigma\|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

Under Assumption 4.4.5 (b), $\lambda_{\min}(\nabla^2 f(x)) \geq \sigma$ for all $x \in \mathbb{R}^n$ and $\Lambda_k = 0$ for all $k \geq 0$.

First, we give an upper bound of $\|d^k(\nu)\|$.

Lemma 4.4.9. *Suppose that Assumption* 4.4.5 *holds. Then,*

$$\|d^k(\nu)\| \leq \frac{1}{\sigma}\|g^k\|.$$

Proof. Based on the proof of Lemma 4.3.1, we get the inequality (4.3.3). Then, we have

$$\|d^k(\nu)\| \leq \frac{1}{\lambda_{\min}(H_k + E_k(\nu))}\|g^k\| \leq \frac{1}{\sigma}\|g^k\|,$$

where the second inequality follows from the fact that $\lambda_{\min}(H_k + E_k(\nu)) \geq \sigma$. □

From the above lemma, we show that $\rho_k(d^k(\nu), \nu) \geq \eta_1$ if ν is greater than a specific value independent of k.

Lemma 4.4.10. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Suppose also that*

$$\nu \geq \max\left(\frac{L_H L_f^{1-\delta}}{2c_2\sigma(1-\eta_1)}, \nu_0\right)$$

Then,

$$\rho_k(d^k(\nu), \nu) \geq \eta_1.$$

Proof. Based on the proof of Lemma 4.4.1, we get the inequality (4.4.2). Then, we have

$$\begin{aligned}
& f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu), \nu)) \\
& \geq \frac{1}{2}\left((1-\eta_1)c_2\nu\|g^k\|^\delta - \frac{L_H}{2}\|d^k(\nu)\|\right)\|d^k(\nu)\|^2 \\
& \geq \frac{1}{2}\left((1-\eta_1)c_2\nu - \frac{L_H\|g^k\|^{1-\delta}}{2\sigma}\right)\|g^k\|^\delta\|d^k(\nu)\|^2 \\
& \geq \frac{1}{2}\left((1-\eta_1)c_2\nu - \frac{L_H L_f^{1-\delta}}{2\sigma}\right)\|g^k\|^\delta\|d^k(\nu)\|^2 \\
& \geq 0,
\end{aligned}$$

where the second inequality follows from Lemma 4.4.9, the third inequality follows from (4.3.2), and the last inequality follows from the assumption on ν. Therefore, by the definition of $\rho_k(d^k(\nu), \nu)$, it implies that $\rho_k(d^k(\nu), \nu) \geq \eta_1$. □

From Lemma 4.4.10, we show that the parameter ν_k^* is bounded from above by some positive constant independent of k.

Lemma 4.4.11. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$\nu_k^* \leq \widehat{U}_\nu,$$

where

$$\widehat{U}_\nu := \gamma_2 \max\left(\frac{L_H L_f^{1-\delta}}{2c_2\sigma(1-\eta_1)}, \nu_0\right).$$

Proof. By using Lemma 4.4.10 and a way similar to the proof of Lemma 4.3.4, we obtain the desired inequality. □

Now, we show that the number l_k^* of inner iterations at the k-th iteration is bounded from above by some positive constant independent of k.

Theorem 4.4.9. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$l_k^* \leq \left\lceil \log_{\gamma_2}\left(\frac{\widehat{U}_\nu}{\nu_{\min}}\right) + 1 \right\rceil.$$

Proof. From Lemma 4.4.11, we have $\nu_{\min} \leq \bar{\nu}_{l_k} \leq \widehat{U}_\nu$. Therefore, by the updating rule of ν, we obtain the desired inequality. □

Next, we show the following lemma.

Lemma 4.4.12. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$\|g^{k+1}\| \leq c_7\|g^k\|^{1+\delta},$$

where

$$c_7 := \frac{c_4}{\sigma} + \frac{L_H L_f^{1-\delta}}{2\sigma^2}.$$

Proof. Based on the proof of Lemma 4.4.6, we get the inequality (4.4.4). Then, we have

$$\begin{aligned}\|g^{k+1}\| &\le c_3\Lambda_k\|d^k(\nu_k^*)\| + c_4\|g^k\|^\delta\|d^k(\nu_k^*)\| + \frac{L_H}{2}\|d^k(\nu_k^*)\|^2\\ &\le \left(\frac{c_4}{\sigma} + \frac{L_H\|g^k\|^{1-\delta}}{2\sigma^2}\right)\|g^k\|^{1+\delta}\\ &\le \left(\frac{c_4}{\sigma} + \frac{L_H L_f^{1-\delta}}{2\sigma^2}\right)\|g^k\|^{1+\delta},\end{aligned}$$

where the second inequality follows from Lemma 4.4.9 and the fact that $\Lambda_k = 0$, and the last inequality follows from (4.3.2). □

Now, we give a lower bound of the reduction $f_k - f_{k+1}$.

Lemma 4.4.13. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$f_k - f_{k+1} \ge p_3\|g^{k+1}\|^{\frac{2}{1+\delta}},$$

where

$$p_3 := \frac{\eta_1}{2c_7^{\frac{2}{1+\delta}}((1+c_3)L_g + c_4\hat{U}_\nu L_f^\delta)}.$$

Proof. Since $\rho_k(d^k(\nu_k^*), \nu_k^*) \ge \eta_1$ from the definition of ν_k^*, we have

$$\begin{aligned}f_k - f_{k+1} &\ge \eta_1(f_k - \phi_k(d^k(\nu_k^*), \nu_k^*))\\ &\ge \frac{\eta_1}{2((1+c_3)L_g + c_4\nu_k^* L_f^\delta)}\|g^k\|^2\\ &\ge \frac{\eta_1}{2c_7^{\frac{2}{1+\delta}}((1+c_3)L_g + c_4\hat{U}_\nu L_f^\delta)}\|g^{k+1}\|^{\frac{2}{1+\delta}},\end{aligned}$$

where the second inequality follows from Lemma 4.3.5, the third inequality follows from Lemmas 4.4.11 and 4.4.12. □

From Lemma 4.4.13, we give an upper bound of K_{outer}.

Theorem 4.4.10. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$K_{\text{outer}} = O(\varepsilon_g^{-\frac{2}{1+\delta}}).$$

Proof. By using Lemma 4.4.13 and a way similar to the proof of Theorem 4.4.2, we obtain the lemma. □

From Theorems 4.4.9 and 4.4.10, we can directly give the following global complexity bound K_{total}.

Theorem 4.4.11. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$K_{\text{total}} = O(\varepsilon_g^{-\frac{2}{1+\delta}}).$$

In particular, if $\delta = 1$*, then*

$$K_{\text{total}} = O(\varepsilon_g^{-1}).$$

By using a technique in [53], we can show that $\{f_k - f(x^*)\}$ converges to 0 linearly.

Theorem 4.4.12. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,* $\{f_k - f(x^*)\}$ *globally linearly converges to* 0*. Thus,*

$$\widehat{K}_{\text{outer}} = O(\log \varepsilon_f^{-1}).$$

Proof. Since f is strongly convex, we have

$$f_{k+1} - f(x^*) \leq g^{k+1^T}(x^{k+1} - x^*) \leq \|g^{k+1}\| \cdot \|x^{k+1} - x^*\| \leq \frac{1}{\sigma}\|g^{k+1}\|^2.$$

It then follows from Lemma 4.4.13 that

$$f_k - f_{k+1} \geq p_3\sigma^{\frac{1}{1+\delta}}\left(f_{k+1} - f(x^*)\right)^{\frac{1}{1+\delta}}.$$

Denoting $\alpha_k := f_k - f(x^*)$ and $\gamma := p_3\sigma^{\frac{1}{1+\delta}}$, we obtain

$$\alpha_k \geq \alpha_{k+1} + \gamma\alpha_{k+1}^{\frac{1}{1+\delta}}.$$

Then, we have from $\alpha_{k+1} \leq \alpha_0$ that

$$\alpha_{k+1} \leq \frac{1}{1 + \gamma\alpha_k^{-\frac{\delta}{1+\delta}}}\alpha_k \leq \frac{1}{1 + \gamma\alpha_0^{-\frac{\delta}{1+\delta}}}\alpha_k. \tag{4.4.20}$$

Therefore, $f_k - f(x^*)$ globally linearly converges to 0.

Next, we show the second part of the theorem. From (4.4.20), we have

$$\alpha_k \leq \left(\frac{1}{1 + \gamma\alpha_0^{-\frac{\delta}{1+\delta}}}\right)^k \alpha_0,$$

and hence if

$$k \geq \frac{1}{\log\left(1 + \gamma\alpha_0^{-\frac{\delta}{1+\delta}}\right)} \log\frac{\alpha_0}{\varepsilon_f},$$

then $\alpha_k \leq \varepsilon_f$. This completes the proof. □

From Theorems 4.4.9 and 4.4.12, we can directly give the following global complexity bound $\widehat{K}_{\text{total}}$.

Theorem 4.4.13. *Suppose that Assumptions* 4.2.1, 4.3.1, 4.4.1 *and* 4.4.5 *hold. Then,*

$$\widehat{K}_{\text{total}} = O(\log \varepsilon_f^{-1}).$$

4.5 Local convergence

In this section, we show that the adaptive RNM converges superlinearly under the local error bound condition. In order to prove the superlinear convergence, we use techniques in Chapter 3, where the regularized Newton method with Armijo's step size rule is shown to have a superlinear rate of convergence under the local error bound condition.

First, we make the following assumptions.

Assumption 4.5.1.

(a) $0 < \delta < 1$.

(b) *There exists a local optimal solution* x^* *of the problem* (4.1.1).

(c) $\nabla^2 f$ *is locally Lipschitz continuous near* x^*, *i.e., there exist constants* $b_1 \in (0,1)$ *and* $\bar{L}_H > 0$ *such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \bar{L}_H \|x - y\|, \quad \forall x, y \in B(x^*, b_1).$$

(d) $\|\nabla f(x)\|$ *provides a local error bound for the problem* (4.1.1) *on* $B(x^*, b_1)$, *i.e., there exists a constant* $c_8 > 0$ *such that*

$$c_8 \mathrm{dist}(x, X^*) \leq \|\nabla f(x)\|, \quad \forall x \in B(x^*, b_1),$$

where X^* *is the local optimal solution set of* (4.1.1).

Note that, under Assumption 4.5.1 (c), the following inequality holds.

$$\|\nabla f(x) - \nabla f(y) - \nabla f(x)(x - y)\| \leq \frac{1}{2}\bar{L}_H \|x - y\|^2, \quad \forall x, y \in B(x^*, b_1). \tag{4.5.1}$$

Moreover, since f is twice continuously differentiable, there exists a positive constant $\bar{L}_g$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq \bar{L}_g \|x - y\|, \quad \forall x, y \in B(x^*, b_1). \tag{4.5.2}$$

In what follows, $\bar{x}^k$ denotes an arbitrary vector such that

$$\|x^k - \bar{x}^k\| = \mathrm{dist}(x^k, X^*), \quad \bar{x}^k \in X^*.$$

First, we show that $\|d^k(\nu)\| = O(\mathrm{dist}(x^k, X^*))$.

Lemma 4.5.1. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Suppose also that* $x^k \in B(x^*, b_1/2)$. *Then,*

$$\|d^k(\nu)\| \leq c_9 \mathrm{dist}(x^k, X^*), \quad \forall \nu \in [\nu_{\min}, \infty),$$

where

$$c_9 := \frac{\bar{L}_H}{2c_2 c_8^\delta \nu_{\min}} + 1 + \frac{\max(c_3, c_4)}{\min(c_1, c_2)}.$$

Proof. Since $x^k \in B(x^*, b_1/2)$, we have $\bar{x}^k \in B(x^*, b_1)$. It then follows from (4.5.1) and the definition of $d^k(\nu)$ that

$$\begin{aligned}
\|d^k(\nu)\| &= \left\|(H_k + E_k(\nu))^{-1} g^k\right\| \\
&= \left\|(H_k + E_k(\nu))^{-1}\Big(g^k - \nabla f(\bar{x}^k) - H_k(x^k - \bar{x}^k) + H_k(x^k - \bar{x}^k)\Big)\right\| \\
&\le \left\|(H_k + E_k(\nu))^{-1}\Big(g^k - \nabla f(\bar{x}^k) - H_k(x^k - \bar{x}^k)\Big)\right\| + \left\|(H_k + E_k(\nu))^{-1} H_k(x^k - \bar{x}^k)\right\| \\
&\le \frac{\bar{L}_H}{2}\|x^k - \bar{x}^k\|^2 \left\|(H_k + E_k(\nu))^{-1}\right\| + \|x^k - \bar{x}^k\| \cdot \left\|(H_k + E_k(\nu))^{-1} H_k\right\| \\
&= \frac{\bar{L}_H}{2}\operatorname{dist}(x^k, X^*)^2 \left\|(H_k + E_k(\nu))^{-1}\right\| + \operatorname{dist}(x^k, X^*) \left\|(H_k + E_k(\nu))^{-1} H_k\right\|, \qquad (4.5.3)
\end{aligned}$$

where the second equality follows from the fact that $\nabla f(\bar{x}^k) = 0$. First, we consider $\|(H_k + E_k(\nu))^{-1}\|$. Then, we have

$$\begin{aligned}
\|(H_k + E_k(\nu))^{-1}\| &= \lambda_{\max}\Big((H_k + E_k(\nu))^{-1}\Big) \\
&= \frac{1}{\lambda_{\min}(H_k + E_k(\nu))} \\
&\le \frac{1}{c_1\Lambda_k + c_2\nu\|g^k\|^\delta} \qquad (4.5.4) \\
&\le \frac{1}{c_2\nu\|g^k\|^\delta} \\
&\le \frac{1}{c_2 c_8^\delta \nu_{\min}}\operatorname{dist}(x^k, X^*)^{-\delta}, \qquad (4.5.5)
\end{aligned}$$

where the first inequality follows from Assumption 4.2.1, and the last inequality follows from Assumption 4.5.1 and the assumption on ν. Next, we consider $\|(H_k + E_k(\nu))^{-1} H_k\|$. Then, we have

$$\begin{aligned}
\|(H_k + E_k(\nu))^{-1} H_k\| &= \|(H_k + E_k(\nu))^{-1}(H_k + E_k(\nu) - E_k(\nu))\| \\
&= \|I - (H_k + E_k(\nu))^{-1} E_k(\nu)\| \\
&\le \|I\| + \|(H_k + E_k(\nu))^{-1}\| \cdot \|E_k(\nu)\| \\
&\le 1 + \frac{c_3\Lambda_k + c_4\nu\|g^k\|^\delta}{c_1\Lambda_k + c_2\nu\|g^k\|^\delta} \\
&\le 1 + \frac{\max(c_3, c_4)}{\min(c_1, c_2)},
\end{aligned}$$

where the second inequality follows from (4.5.4) and Assumption 4.2.1. It then follows from (4.5.3) and (4.5.5) that

$$\|d^k(\nu)\| \le \frac{\bar{L}_H}{2c_2 c_8^\delta \nu_{\min}}\operatorname{dist}(x^k, X^*)^{2-\delta} + \left(1 + \frac{\max(c_3, c_4)}{\min(c_1, c_2)}\right)\operatorname{dist}(x^k, X^*). \qquad (4.5.6)$$

Since $x^k \in B(x^*, b_1/2)$, we have $\operatorname{dist}(x^k, X^*) \le 1$. Thus, $\delta < 1$ implies that

$$\operatorname{dist}(x^k, X^*)^{2-\delta} \le \operatorname{dist}(x^k, X^*).$$

It then follows from (4.5.6) that

$$\|d^k(\nu)\| \leq \left(\frac{\bar{L}_H}{2c_2c_8^\delta\nu_{\min}} + 1 + \frac{\max(c_3, c_4)}{\min(c_1, c_2)} \right) \mathrm{dist}(x^k, X^*),$$

which is the desired inequality. □

From Lemma 4.5.1, we can show that $x^k + d^k(\nu) \in B(x^*, b_1)$ if x^k is sufficiently close to x^*.

Lemma 4.5.2. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Let* $b_2 := b_1/(c_9+1)$. *Suppose also that* $x^k \in B(x^*, b_2)$. *Then,*

$$x^k + d^k(\nu) \in B(x^*, b_1), \quad \forall \nu \in [\nu_{\min}, \infty),$$

Proof. Since $b_2 \leq b_1/2$, we have $x^k \in B(x^*, b_1/2)$. It then follows from Lemma 4.5.1 that

$$\begin{aligned} \|x^k + d^k(\nu) - x^*\| &\leq \|x^k - x^*\| + \|d^k(\nu)\| \leq \|x^k - x^*\| + c_9 \mathrm{dist}(x^k, X^*) \\ &\leq \|x^k - x^*\| + c_9\|x^k - x^*\| \leq b_1, \end{aligned}$$

where the last inequality follows from the fact that $x^k \in B(x^*, b_2)$. □

From Lemma 4.5.2 and the convexity of the set $B(x^*, b_1)$, we have

$$x^k + \tau d^k(\nu) \in B(x^*, b_1), \quad \forall \tau \in [0, 1], \quad \forall \nu \in [\nu_{\min}, \infty)$$

if $x^k \in B(x^*, b_2)$. It then follows from Assumption 4.5.1 (c) that

$$\|\nabla^2 f(x^k + \tau d^k(\nu)) - H_k\| \leq \bar{L}_H \tau \|d^k(\nu)\|, \quad \forall \tau \in [0, 1], \quad \forall \nu \in [\nu_{\min}, \infty). \tag{4.5.7}$$

Now, we show that $l_k^* = 1$ and $\nu_k^* \leq \nu_{k-1}^*$ if x^k is sufficiently close to x^*.

Lemma 4.5.3. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Let*

$$b_3 := \min\left(b_2, \left(\frac{2(1-\eta_1)c_2c_8^\delta\nu_{\min}}{c_9\bar{L}_H} \right)^{\frac{1}{1-\delta}} \right).$$

Suppose also that $x^k \in B(x^*, b_3)$. *Then,*

$$l_k^* = 1, \quad \nu_k^* \leq \nu_{k-1}^*.$$

In particular, if $\{x^0, x^1, \ldots, x^k\} \subseteq B(x^*, b_3)$, *then*

$$\nu_k^* \leq \nu_0.$$

Proof. From Lemma 4.2.4 and (4.5.7), we have

$$\begin{aligned}
&f_k - f(x^k + d^k(\nu)) - \eta_1(f_k - \phi_k(d^k(\nu),\nu)) \\
&\geq \frac{1}{2}\left((1-\eta_1)c_2\nu\|g^k\|^\delta - \bar{L}_H\|d^k(\nu)\|\int_0^1 \tau \mathrm{d}\tau\right)\|d^k(\nu)\|^2 \\
&= \frac{1}{2}\left((1-\eta_1)c_2\nu\|g^k\|^\delta - \frac{\bar{L}_H}{2}\|d^k(\nu)\|\right)\|d^k(\nu)\|^2 \\
&\geq \frac{1}{2}\left((1-\eta_1)c_2c_8^\delta\nu - \frac{c_9\bar{L}_H}{2}\mathrm{dist}(x^k,X^*)^{1-\delta}\right)\mathrm{dist}(x^k,X^*)^\delta\|d^k(\nu)\|^2 \\
&\geq \frac{1}{2}\left((1-\eta_1)c_2c_8^\delta\nu - \frac{c_9\bar{L}_H}{2}\|x^k - x^*\|^{1-\delta}\right)\mathrm{dist}(x^k,X^*)^\delta\|d^k(\nu)\|^2 \\
&\geq \frac{(1-\eta_1)c_2c_8^\delta(\nu-\nu_{\min})}{2}\mathrm{dist}(x^k,X^*)^\delta\|d^k(\nu)\|^2,
\end{aligned}$$

where the second inequality follows from Assumption 4.5.1 (d) and Lemma 4.5.1, and the last inequality follows from the fact that $x^k \in B(x^*, b_3)$. Then, for any $\nu \geq \nu_{\min}$, we have $\rho_k(d^k(\nu),\nu) \geq \eta_1$, and hence $l_k^* = 1$ and $\nu_k^* \leq \nu_{k-1}^*$. The second part of the Lemma directly follows from the updating rule of ν. □

Next, we give the following lemma.

Lemma 4.5.4. *Suppose that Assumption* 4.5.1 *holds. Then,*

$$\Lambda_k \leq \bar{L}_H \mathrm{dist}(x^k, X^*).$$

Proof. In a way similar to the proof of Lemma 3.5.2, we can show the lemma. □

By using Lemma 4.5.4, we show that $\mathrm{dist}(x^k, X^*)$ converges to 0 superlinearly, as long as $\{x^k\}$ lies in a neighborhood of x^*.

Lemma 4.5.5. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Suppose also that* $\{x^0, x^1, \ldots, x^k, x^{k+1}\} \subseteq B(x^*, b_3)$. *Then,*

$$\mathrm{dist}(x^{k+1}, X^*) = O\left(\mathrm{dist}(x^k, X^*)^{1+\delta}\right).$$

Proof. From Assumption 4.5.1 (d) that

$$\begin{aligned}
\mathrm{dist}(x^{k+1}, X^*) &\leq \frac{1}{c_8}\|g^{k+1}\| \\
&\leq \frac{1}{c_8}\|-H_k d^k(\nu_k^*) - g^k\| + \frac{\bar{L}_H}{2c_8}\|d^k(\nu_k^*)\|^2 \\
&= \frac{1}{c_8}\|E_k(\nu_k^*)d^k(\nu_k^*)\| + \frac{\bar{L}_H}{2c_8}\|d^k(\nu_k^*)\|^2 \\
&\leq \frac{c_3}{c_8}\Lambda_k\|d^k(\nu_k^*)\| + \frac{c_4\nu_k^*}{c_8}\|g^k\|^\delta\|d^k(\nu_k^*)\| + \frac{\bar{L}_H}{2c_8}\|d^k(\nu_k^*)\|^2 \\
&\leq \frac{c_3}{c_8}\Lambda_k\|d^k(\nu_k^*)\| + \frac{c_4\nu_0}{c_8}\|g^k\|^\delta\|d^k(\nu_k^*)\| + \frac{\bar{L}_H}{2c_8}\|d^k(\nu_k^*)\|^2, \qquad (4.5.8)
\end{aligned}$$

where the second inequality follows from (4.5.1), the third inequality follows from Assumption 4.2.1 (c), and the last inequality follows from Lemma 4.5.3. From (4.5.2) and the fact that $\nabla f(\bar{x}^k) = 0$, we have

$$\|g^k\|^\delta = \|g^k - \nabla f(\bar{x}^k)\|^\delta \leq \bar{L}_g^\delta \mathrm{dist}(x^k, X^*)^\delta.$$

It then follows from (4.5.8) and Lemmas 4.5.1 and 4.5.4 that

$$\begin{aligned}
\mathrm{dist}(x^{k+1}, X^*) &\leq \frac{c_3 c_9^2 \bar{L}_H}{c_8}\mathrm{dist}(x^k, X^*)^2 + \frac{c_4 c_9 \nu_0 \bar{L}_g^\delta}{c_8}\mathrm{dist}(x^k, X^*)^{1+\delta} + \frac{c_9^2 \bar{L}_H}{2c_8}\mathrm{dist}(x^k, X^*)^2 \\
&\leq \frac{c_9(2c_3 c_9 \bar{L}_H + 2c_4 \nu_0 \bar{L}_g^\delta + c_9 \bar{L}_H)}{2c_8}\mathrm{dist}(x^k, X^*)^{1+\delta},
\end{aligned}$$

where the last inequality follows from $\mathrm{dist}(x^k, X^*) \leq 1$. □

From Lemma 4.5.5, there exists a positive constant $b_4 \leq b_3$ such that

$$\mathrm{dist}(x^k, X^*) \leq b_4 \Rightarrow \mathrm{dist}(x^{k+1}, X^*) \leq \frac{1}{2}\mathrm{dist}(x^k, X^*).$$

Note that, Lemma 4.5.5 shows that $\{\mathrm{dist}(x^k, X^*)\}$ converges to 0 superlinearly if $x^k \in B(x^*, b_3)$ for all k. Now we give a sufficient condition for $x^k \in B(x^*, b_3)$ for all k.

Lemma 4.5.6. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Let* $b_5 := \min(b_3, b_4)$ *and* $b_6 := \frac{1}{1+2c_9}b_5$. *Suppose also that* $x^0 \in B(x^*, b_6)$. *Then,* $x^k \in B(x^*, b_5)$ *for all* k.

Proof. In a way similar to the proof of [68, Lemma 2.3], we obtain the desired result. □

By using Lemmas 4.5.5 and 4.5.6, we give the rate of convergence.

Theorem 4.5.1. *Suppose that Assumptions* 4.2.1 *and* 4.5.1 *hold. Suppose also that* $x^0 \in B(x^*, b_6)$. *Then,* $\{\mathrm{dist}(x^k, X^*)\}$ *converges to* 0 *at the rate of* $1+\delta$. *Moreover,* $\{x^k\}$ *converges to a local optimal solution* $\hat{x} \in B(x^*, b_5)$.

Proof. The first part of the theorem directly follows from Lemmas 4.5.5 and 4.5.6. Moreover, in a way similar to the proof of [68, Theorem 2.1], we obtain the result of the second part of the theorem. □

Remark 4.5.1. *Note that in a way similar to the proof of* [38, Theorem 3.2], *we can prove that* $\{x^k\}$ *converges to* $\hat{x}$ *at the rate of* $1 + \delta$.

Remark 4.5.2. *We get a rapid local convergence if we take a larger* δ. *However, we cannot guarantee the quadratic convergence since* δ *must be less than* 1 *from Lemma* 4.5.3. *Note that, when the second-order sufficient optimality condition holds at* x^*, *we can prove that the adaptive RNM with* $\delta = 1$ *has quadratic convergence as well as the extended RNM in Chapter* 3.

4.6 Numerical results

In this section, we report some results on the following numerical experiments for the adaptive RNM.

1. Examination of the effects of the regularization matrix.
2. Comparison of the adaptive RNM and the RNM using Armijo's line search.

In each experiment, benchmark problems were chosen from CUTEr [28]. All algorithms were coded in MATLAB 7.4, and run on a machine with 3.2GHz Pentium 4 CPU and 3.2GB memory. We used an initial point x^0 given by CUTEr, and set the termination criterion as $\|g^k\| \leq 10^{-5}$. If the number of inner iterations at the k-th iteration or the number of outer iterations exceeds 10^4, then we terminated all methods as failing.

We consider the following two regularization matrix $E_k(\nu)$.

(A) $E_k(\nu) = (c\Lambda_k + \nu\|g^k\|^\delta)I$.

(B) $E_k(\nu) = (c\Lambda_k + \nu\min(1, \|g^k\|^\delta))I$.

The updating rule (B) prevents $\|d^k(\bar{\nu}_{l_k})\|$ from becoming too small when $\|g^k\|^\delta$ is large. Note that, the convergence properties given in Sections 4.3 – 4.5 still hold even if we use the matrix (B). We updated ν in Steps 2 and 3 as follows.

$$\begin{aligned} \rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1 &\Rightarrow \bar{\nu}_{l_k+1} = \gamma_b\bar{\nu}_{l_k}, \\ \eta_2 > \rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1 &\Rightarrow \nu_{k+1} = \bar{\nu}_{l_k}, \\ \rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_2 &\Rightarrow \nu_{k+1} = \max\left(\nu_{\min}, \gamma_a\bar{\nu}_{l_k}\right), \end{aligned}$$

where γ_a and γ_b are positive parameters such that $\gamma_a < 1$ and $\gamma_b > 1$. In all numerical experiments, except for γ_a, γ_b and δ, the parameters of the adaptive RNM are chosen as follows.

$$\nu_0 = 1,\ \nu_{\min} = 10^{-5},\ c = 2,\ \eta_1 = 0.01,\ \eta_2 = 0.8.$$

In Subsections 4.6.1 and 4.6.2, we will compare algorithms by using the distribution function proposed in [21]. We denote a set of solvers as S, and a set of problems that can be solved by all methods in S as P_S. We also denote a measure for evaluation required to solve a problem p by a solver s as $t_{p,s}$, and the best $t_{p,s}$ for each p as t_p^*, i.e., $t_p^* := \min\{t_{p,s} \mid s \in S\}$. The distribution function $F_s^S(\tau)$ for a method s is defined by

$$F_s^S(\tau) = \frac{|\{p \in P_S \mid t_{p,s} \leq \tau t_p^*\}|}{|P_S|}, \quad \tau \geq 1.$$

4.6.1 Influences of the regularization matrix

First, we investigate influences of the parameter δ and the regularization matrix (A) and (B). We set γ_a and γ_b as $\gamma_a = 0.5$ and $\gamma_b = 2$, respectively.

Figure 4.1 shows the distribution functions for the adaptive RNM with various δ and the matrix (A) and (B) in terms of the number of function evaluations. Figure 4.1 shows that for $\delta = 0.5$, the matrix (A) is the almost same as the matrix (B). On the other hand, for $\delta = 1$ and 2, the matrix (B) is better than the matrix (A). The reason is that when $\|g^k\|^\delta$ is large, $\|d^k(\bar{\nu}_{l_k})\|$ becomes too small, and a sequence of the adaptive RNM changes only slightly. Moreover, from the same reason, the number of function evaluations tends to become large as δ become large for the matrix (A). Finally, for the matrix (B), the adaptive RNM does not have much difference among $\delta = 0.5, 1, 2$. From the above fact, the adaptive RNM has good numerical performance when we use the matrix (B).

Next, we examine the influences of (γ_a, γ_b). We set $\delta = 1$ and used the matrix (B), and tested the adaptive RNM for each (γ_a, γ_b) in $\{\frac{1}{2}, \frac{1}{10}\} \times \{2, 10\}$.

Figure 4.2 shows the comparisons of (γ_a, γ_b) in terms of the number of function evaluations. From Figure 4.2, we see that $\gamma_b = 10$ has good performances as compared to $\gamma_b = 2$.

4.6.2 Comparison with the regularized Newton method using line search

We compare the adaptive RNM with the extended RNM, which is proposed in Chapter 3 and uses Armijo's step size rule. In the adaptive RNM, we adopted the updating rule (B) of μ_k, and set $\delta = 1$, $\gamma_a = 1/10$ and $\gamma_b = 10$. For comparison, we used the following extended RNM.

The extended regularized Newton method

Step 0 : Choose a starting point x^0. Set $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : Compute

$$d^k := -(H_k + (2\Lambda_k + \min(1, \|g^k\|))I)^{-1} g^k.$$

Step 3 : Find the smallest nonnegative integer l_k such that

$$f_k - f(x^k + (0.5)^{l_k} d^k) \geq -0.01 \times (0.5)^{l_k} {g^k}^T d^k.$$

Step 4 : Set $x^{k+1} := x^k + (0.5)^{l_k} d^k$ and $k := k + 1$. Go to Step 1.

Table A.1 shows the number of function evaluations N_f and the number of solving linear equations N_L for each method. Note that, the computational complexity of calculating the minimum eigenvalue of H_k is not contained in N_L. The symbol "$-$" in the table means that the number of inner or outer iterations of the solution methods exceed 10^4. The adaptive RNM cannot solve 'MARATOSB'.

Figures 4.3 and 4.4 show the comparisons of the adaptive RNM and the extended RNM for N_f and N_L. From Figures 4.3 and 4.4, we see that both N_f and N_L of the adaptive RNM are much less than those of the extended RNM.

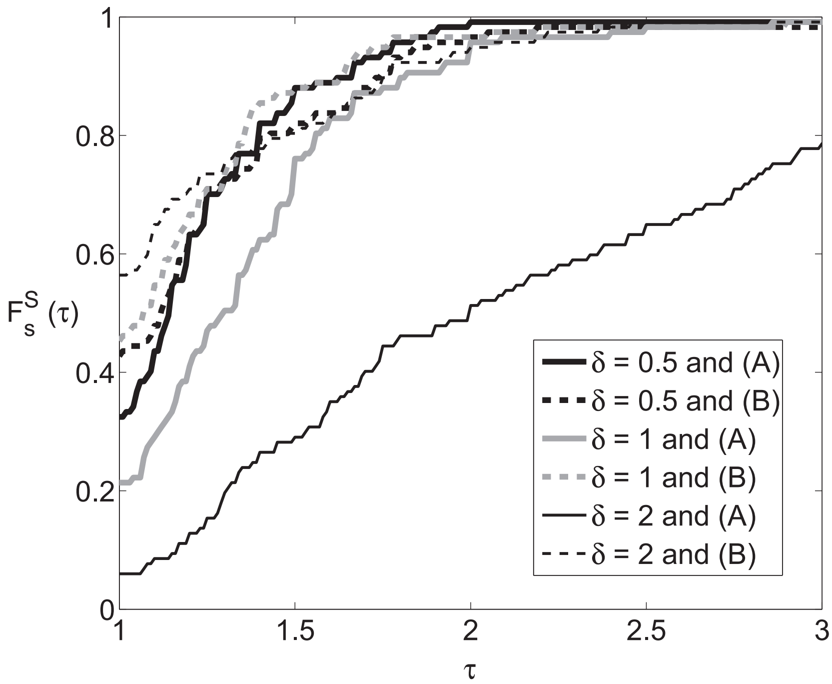

Figure 4.1: Comparison among different values of δ in (A) and (B)

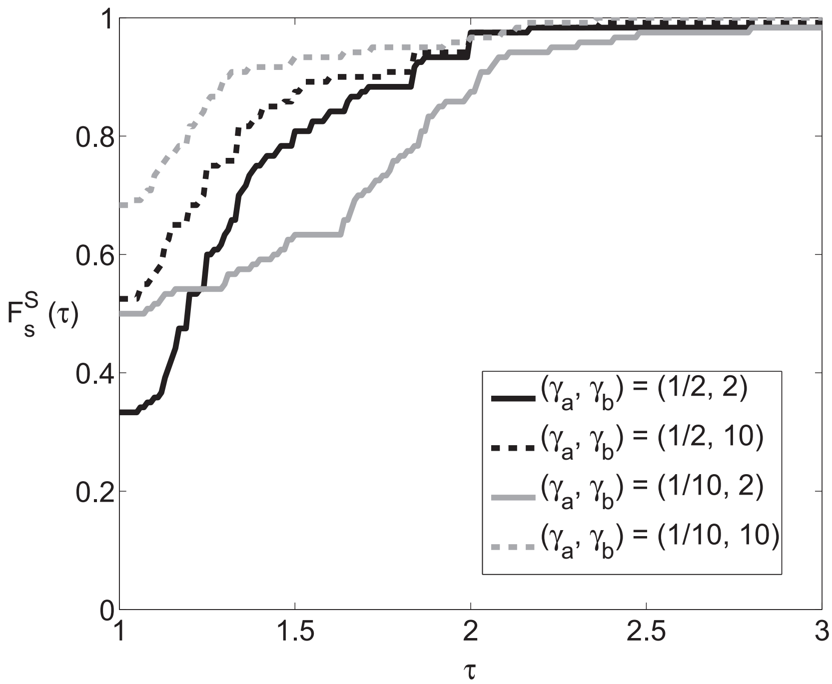

Figure 4.2: Comparison among different pairs of (γ_a, γ_b)

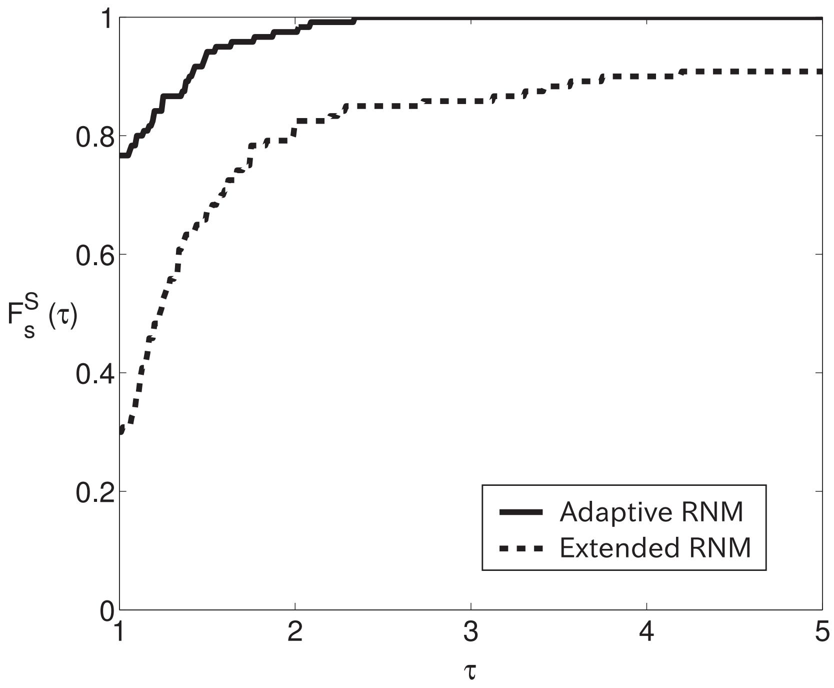

Figure 4.3: Comparison of the adaptive RNM and the extended RNM for N_f

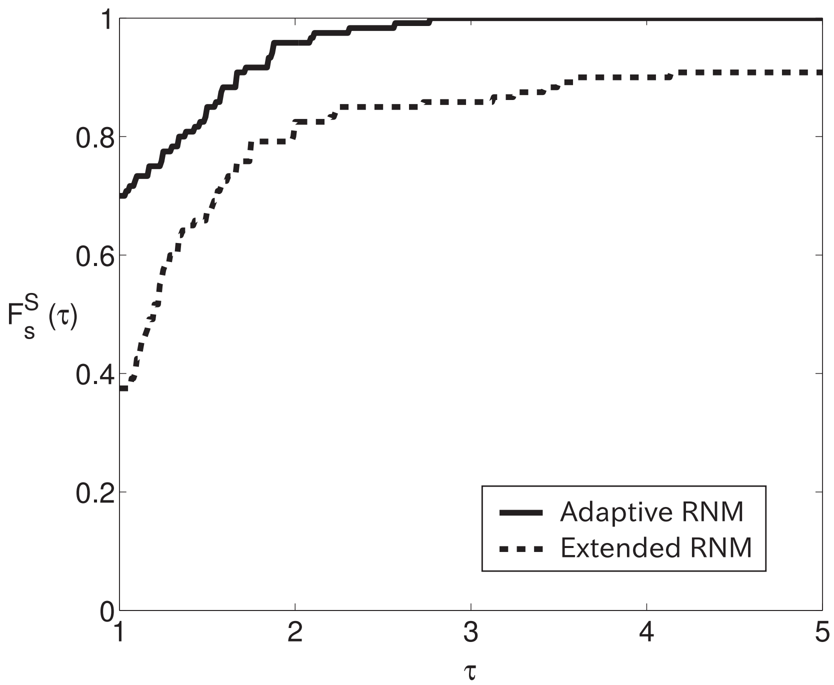

Figure 4.4: Comparison of the adaptive RNM and the extended RNM for N_L

4.7 Concluding remarks

In this chapter, we have proposed an RNM without line search (called adaptive RNM). We have shown the global convergence and the superlinear convergence of the adaptive RNM, and given its global complexity bounds. Moreover, we have presented some numerical results, which show that the number of function evaluations and the number of solving linear equations of the adaptive RNM are much less than those of the RNM using Armijo's step size rule in many benchmark problems.

In Section 4.6, we adopted the simple regularization matrix $E_k(\nu)$. As described in Section 4.2, we can compute a search direction by the modified Cholesky factorization without calculating the minimum eigenvalue of H_k. Then, the adaptive RNM may have better performance in numerical experiments.

Chapter 5

Global complexity bound analysis of the Levenberg-Marquardt method for the nonlinear least squares problem with a smooth mapping

5.1 Introduction

In this chapter and the next chapter, we consider the following nonlinear least squares problem.

$$\text{minimize } \ f(x) := \frac{1}{2}\|F(x)\|^2, \tag{5.1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuous mapping. Here we assume that F is continuously differentiable. In the next chapter, we will consider the case where F is nonsmooth.

The Levenberg-Marquardt method (LMM) is one of the solution methods for (5.1.1) [5, 45, 54, 55, 67]. For a current point x^k, the LMM adopts a search direction $d^k(\mu_k)$ given by

$$d^k(\mu_k) := -(J_k^T J_k + \mu_k I)^{-1} J_k^T F^k, \tag{5.1.2}$$

where μ_k is a positive parameter. Taking $\mu_k \to \infty$, we have $\|d^k(\mu_k)\| \to 0$ and $d^k(\mu_k)/\|d^k(\mu_k)\| \to J_k^T F^k/\|J_k^T F^k\|$. Therefore, $f(x^k + d^k(\mu_k)) < f_k$ for μ_k sufficiently large, and hence the LMM converges globally to a stationary point of f if μ_k is appropriately updated. To guarantee a global convergence of the LMM, many updating rules of the regularization parameter μ_k, such as Osborne's rule [55] and Moré's rule [45], have been proposed. In this chapter, we adopts Osborne's rule since a search direction $d^k(\mu_k)$ is given as a solution of the linear equations, which is much easier to solve than the trust-region subproblem in Moré's rule. The details of the LMM with Osborne's rule is presented in Section 5.2.

As described in Section 1.4, global complexity bounds have been vigorously discussed for solution methods of unconstrained minimization problems. Until now, some bounds for the steepest descent method and the Newton-type methods have been presented [6, 9, 29, 50, 53, 57]. Thus, by applying these methods to the problem (5.1.1), we can give the global complexity bound, and estimate the worst computational time in advance. However, since these method are not specialized to nonlinear least squares problems, they are not efficient. In fact, the Newton-type methods require the second derivative of F, and have to solve nonconvex subproblems at each iteration. Moreover, although the steepest descent method requires only the Jacobian of F, its convergence is slow in general. Thus, it is worth investigating a global complexity bound for methods specified for (5.1.1). Recently, Nesterov [51] proposed a modified Gauss-Newton method for solving a system of nonlinear equations and gave interesting results. However, the modified Gauss-Newton method also has to solve computationally expensive subproblems. The LMM is the special method for (5.1.1), and its global complexity bound remains unknown.

In this chapter, we investigate the global complexity bound of the LMM with Osborne's simple updating rule of the regularization parameter parameter. In particular, we show that the global complexity bound is $O(\varepsilon_g^{-2})$ without assuming the nonsingularity of $\nabla F(x)^T \nabla F(x)$.

5.2 Global complexity bound of the Levenberg-Marquardt method

First, we explain the updating rule of the regularization parameter μ_k proposed by Osborne [55]. Then, we give a global complexity bound of the LMM with the rule. In what follows, we denote the LMM with Osborne's updating rule as the LMM for simplicity.

The LMM adopts a search direction $d^k(\mu_k)$ defined by (5.1.2), and controls μ_k directly as follows. Let $\phi_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a model function of f at x^k defined by

$$\phi_k(d, \mu) := \frac{1}{2}\|F^k + J_k d\|^2 + \frac{1}{2}\mu\|d\|^2.$$

Note that, $d^k(\mu_k)$ is a global minimizer of $\phi_k(\cdot, \mu_k)$. Let $\rho_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the ratio of the reduction of the objective function value to that of the model function value, i.e.,

$$\rho_k(d, \mu) := \frac{f_k - f(x^k + d)}{f_k - \phi_k(d, \mu)}.$$

If $\rho_k(d^k(\mu_k), \mu_k)$ is large, then we adopt $d^k(\mu_k)$ and decrease the parameter μ_k. On the other hand, if $\rho_k(d^k(\mu_k), \mu_k)$ is small, then we increase μ_k and compute $d^k(\mu_k)$ once again.

The precise description of the LMM is as follows.

The Levenberg-Marquardt Method

Step 0 : Choose parameters $\varepsilon_g, \mu_0, \gamma_1, \gamma_2, \eta_1, \eta_2$ such that

$$0 < \varepsilon_g < 1,\ \mu_0 > 0,\ \gamma_1 < 1 < \gamma_2,\ 0 < \eta_1 \le \eta_2 \le 1.$$

Choose a starting point x^0. Set $k := 0$.

Step 1 : If $\|J_k^T F^k\| \le \varepsilon_g$, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\mu}_{l_k} := \mu_k$.

Step 2.1 : Compute $d^k(\bar{\mu}_{l_k})$.

Step 2.2 : Compute $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k})$. If $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) < \eta_1$, then set $\bar{\mu}_{l_k+1} := \gamma_2 \bar{\mu}_{l_k}$ and $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $\eta_2 > \rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq \eta_1$, then set $\mu_{k+1} := \bar{\mu}_{l_k}$. If $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq \eta_2$, then set $\mu_{k+1} := \gamma_1 \bar{\mu}_{l_k}$. Set $x^{k+1} := x^k + d^k(\bar{\mu}_{l_k})$ and $k := k + 1$. Go to Step 1.

Osborne [55] showed that the LMM has a global convergence property under appropriate conditions.

Next, we discuss the global complexity bound of the LMM. In what follows, for simplicity, we denote l_k and $\bar{\mu}_{l_k}$ of the last iteration in the inner loops of Steps 2.0–2.2 at each k as l_k^* and μ_k^*, respectively.

Let K_{outer} be the total number of outer iterations when the algorithm terminates. If there does not exist such K_{outer}, we define $K_{\text{outer}} := \infty$. Moreover, let K_{total} be the total number of inner iterations until $k = K_{\text{outer}}$ holds, i.e.,

$$K_{\text{total}} := \sum_{k=0}^{K_{\text{outer}}-1} l_k^*.$$

Note that, we cannot estimate the total computational time from K_{outer}. In contrast, K_{total} means the total number of solving linear equations. Therefore, the main task of the chapter is to investigate K_{total}. To this end, we firstly make the following assumption.

Assumption 5.2.1. *The level set of f at the initial point x^0 is compact, i.e., $\Omega := \{x \in \mathbb{R}^n \mid f(x) \leq f_0\}$ is compact.*

Since $\{f_k\}$ is monotonically decreasing, the sequence $\{x^k\}$ is included in the compact set Ω. Thus, since F is continuously differentiable, there exists a positive constant $F_{\max}$ such that

$$\|F(x)\| \leq F_{\max}, \quad \forall x \in \Omega. \tag{5.2.1}$$

Moreover, since F is continuously differentiable, F is Lipschitz continuous on Ω, and hence there exists a positive constant L_F such that

$$\|\nabla F(x)\| \leq L_F, \quad \forall x \in \Omega. \tag{5.2.2}$$

Now, we give bounds of eigenvalues of $(J_k^T J_k + \mu I)^{-1}$.

Lemma 5.2.1. *Suppose that Assumption* 5.2.1 *holds. Then, for any $\mu \in (0, \infty)$,*

$$\lambda_{\max}((J_k^T J_k + \mu I)^{-1}) \leq \frac{1}{\mu},$$
$$\lambda_{\min}((J_k^T J_k + \mu I)^{-1}) \geq \frac{1}{L_F^2 + \mu}.$$

Proof. Since $J_k^T J_k$ is positive semidefinite, we have

$$\lambda_{\max}((J_k^T J_k + \mu I)^{-1}) = \frac{1}{\lambda_{\min}(J_k^T J_k + \mu I)} \le \frac{1}{\mu}.$$

On the other hand, from (5.2.2), we have

$$\lambda_{\min}((J_k^T J_k + \mu I)^{-1}) = \frac{1}{\lambda_{\max}(J_k^T J_k + \mu I)} = \frac{1}{\|J_k\|^2 + \mu} \ge \frac{1}{L_F^2 + \mu}.$$

This completes the proof. □

The next lemma indicates that $\|d^k(\mu)\|$ is bounded from above when $\mu \in [\mu_0, \infty)$.

Lemma 5.2.2. *Suppose that Assumption* 5.2.1 *holds. Then, for any* $\mu \in [\mu_0, \infty)$,

$$\|d^k(\mu)\| \le U_d,$$

where

$$U_d := \frac{F_{\max} L_F}{\mu_0}.$$

Proof. From the definition of $d^k(\mu)$, we have

$$\begin{aligned}\|d^k(\mu)\| &= \|(J_k^T J_k + \mu I)^{-1} J_k^T F^k\| \le \|(J_k^T J_k + \mu I)^{-1}\| \cdot \|J_k^T\| \cdot \|F^k\| \\ &\le L_F F_{\max} \lambda_{\max}\Big((J_k^T J_k + \mu I)^{-1}\Big) \le \frac{F_{\max} L_F}{\mu_0},\end{aligned}$$

where the second inequality follows from (5.2.1) and (5.2.2), and the last inequality follows from Lemma 5.2.1 and $\mu \ge \mu_0$. □

In what follows, we further assume the Lipschitz continuity of ∇F.

Assumption 5.2.2. ∇F *is Lipschitz continuous on* $\Omega + B(0, U_d)$ *with modulus* L_J.

Note that, if F is twice continuously differentiable, then Assumption 5.2.2 holds.

Next, we show that $\rho_k(d^k(\mu), \mu) \ge 1$ if μ is greater than a specific value independent of k.

Lemma 5.2.3. *Suppose that Assumptions* 5.2.1 *and* 5.2.2 *hold. Suppose also that*

$$\mu \ge \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3} L_J^2 U_d^2, \mu_0\right).$$

Then,

$$\rho_k(d^k(\mu), \mu) \ge 1.$$

Proof. Since F is continuously differentiable, we have

$$\begin{aligned} F(x^k + d^k(\mu)) &= F^k + \int_0^1 \nabla F(x^k + \tau d^k(\mu)) d^k(\mu) \mathrm{d}\tau \\ &= F^k + J_k d^k(\mu) + \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau. \end{aligned}$$

It then follows from the definitions of f that

$$\begin{aligned} & f(x^k + d^k(\mu)) \\ &= \frac{1}{2} \left\| F^k + J_k d^k(\mu) + \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \right\|^2 \\ &= \frac{1}{2} \left\| F^k + J_k d^k(\mu) \right\|^2 + (F^k + J_k d^k(\mu))^T \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \\ &\quad + \frac{1}{2} \left\| \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \right\|^2 \\ &= \phi_k(d^k(\mu), \mu) + (F^k + J_k d^k(\mu))^T \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \\ &\quad + \frac{1}{2} \left\| \int_0^1 (\nabla F(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \right\|^2 - \frac{1}{2} \mu \|d^k(\mu)\|^2 \\ &\leq \phi_k(d^k(\mu), \mu) + \|F^k + J_k d^k(\mu))\| \cdot \|d^k(\mu)\| \int_0^1 \|\nabla F(x^k + \tau d^k(\mu)) - J_k\| \mathrm{d}\tau \\ &\quad + \frac{1}{2} \|d^k(\mu)\|^2 \int_0^1 \|\nabla F(x^k + \tau d^k(\mu)) - J_k\|^2 \mathrm{d}\tau - \frac{1}{2} \mu \|d^k(\mu)\|^2, \end{aligned} \tag{5.2.3}$$

where the third equality follows from the definition of $\phi_k(d^k(\mu), \mu)$. From the assumption on μ, we have $\mu \geq \mu_0$. Thus, from Lemma 5.2.2 and the fact that the sequence $\{x^k\}$ is included in the compact set Ω, we have $x + \tau d^k(\mu) \in \Omega + B(0, U_d)$ for all $\tau \in [0, 1]$. Hence, from Assumption 5.2.2, we have

$$\|\nabla F(x^k + \tau d^k(\mu)) - J_k\| \leq \tau L_J \|d^k(\mu)\|, \quad \forall \tau \in [0, 1].$$

It then follows from (5.2.3) that

$$\begin{aligned} f(x^k + d^k(\mu)) &\leq \phi_k(d^k(\mu), \mu) + L_J \|F^k + J_k d^k(\mu))\| \cdot \|d^k(\mu)\|^2 \int_0^1 \tau \mathrm{d}\tau \\ &\quad + \frac{1}{2} L_J^2 \|d^k(\mu)\|^4 \int_0^1 \tau^2 \mathrm{d}\tau - \frac{1}{2} \mu \|d^k(\mu)\|^2 \\ &= \phi_k(d^k(\mu), \mu) + \frac{1}{2} \|d^k(\mu)\|^2 \left(L_J \|F^k + J_k d^k(\mu))\| + \frac{1}{3} L_J^2 \|d^k(\mu)\|^2 - \mu \right) \\ &\leq \phi_k(d^k(\mu), \mu) + \frac{1}{2} \|d^k(\mu)\|^2 \left(L_J (F_{\max} + L_F U_d) + \frac{1}{3} L_J^2 U_d^2 - \mu \right) \\ &\leq \phi_k(d^k(\mu), \mu), \end{aligned}$$

where the second inequality follows from (5.2.1), (5.2.2) and Lemma 5.2.2, and the last inequality follows from the assumption on μ. Therefore, by the definition of $\rho_k(d^k(\mu), \mu)$, we have

$$\rho_k(d^k(\mu), \mu) = \frac{f_k - f(x^k + d^k(\mu))}{f_k - \phi_k(d^k(\mu), \mu)} \geq 1,$$

which is the desired inequality. □

From Lemma 5.2.3, we can show that the regularization parameter μ_k^* is bounded from above.

Lemma 5.2.4. *Suppose that Assumptions* 5.2.1 *and* 5.2.2 *hold. Then,*

$$\mu_k^* \leq U_\mu,$$

where

$$U_\mu := \gamma_2 \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right).$$

Proof. From Lemma 5.2.3, if

$$\bar{\mu}_{l_k} \geq \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right),$$

then, $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq 1$, and hence the inner loops of Step 2 must terminate. Therefore, if

$$\bar{\mu}_1 \geq \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right),$$

at the k-th iteration, then $\mu_k^* = \bar{\mu}_1$. On the other hand, if

$$\bar{\mu}_1 < \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right),$$

then μ_k^* must satisfy

$$\mu_k^* \leq \gamma_2 \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right).$$

Otherwise,

$$\bar{\mu}_{l_k^*-1} \geq \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right),$$

which contradicts $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) < \eta_1 \leq 1$. Consequently, we have

$$\begin{aligned}
\mu_k^* &\leq \max\left(\bar{\mu}_1, \gamma_2\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2\right), \gamma_2\mu_0\right) \\
&= \max\left(\mu_{k-1}^*, \gamma_2\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2\right), \gamma_2\mu_0\right) \\
&\leq \cdots \leq \max\left(\mu_0, \gamma_2\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2\right), \gamma_2\mu_0\right) \\
&= \gamma_2 \max\left(L_J(F_{\max} + L_F U_d) + \frac{1}{3}L_J^2 U_d^2, \mu_0\right)
\end{aligned}$$

from the updating rule of μ. □

By using the above lemma, we give a lower bound of the reduction of the objective function when $k < K_{\text{outer}}$.

Lemma 5.2.5. *Suppose that Assumptions* 5.2.1 *and* 5.2.2 *hold. Suppose also that* $k < K_{\text{outer}}$. *Then,*

$$f_k - f_{k+1} > p\varepsilon_g^2.$$

where

$$p := \frac{\eta_1}{2(L_F^2 + U_\mu)}.$$

Proof. First note that $\|J_k^T F^k\| > \varepsilon_g$ for $k < K_{\text{outer}}$. From Lemmas 5.2.1 and 5.2.4, we have

$$\lambda_{\min}\Big((J_k^T J_k + \mu_k^* I)^{-1}\Big) \geq \frac{1}{U_J^2 + U_\mu}. \tag{5.2.4}$$

Since $\rho_k(d^k(\mu_k^*), \mu_k^*) \geq \eta_1$ from the definition of μ_k^*, we have

$$f_k - f_{k+1} \geq \eta_1 (f_k - \phi_k(d^k(\mu_k^*), \mu_k^*)).$$

It then follows from the definitions of f and ϕ_k that

$$\begin{aligned}
f_k - f_{k+1} &\geq \frac{1}{2}\eta_1(\|F^k\|^2 - \|F^k + J_k d^k(\mu_k^*)\|^2 - \mu_k^* \|d^k(\mu_k^*)\|^2) \\
&= \frac{1}{2}\eta_1(-2{F^k}^T J_k d^k(\mu_k^*) - d^k(\mu_k^*)^T (J_k^T J_k + \mu_k^* I) d^k(\mu_k^*)) \\
&= \frac{1}{2}\eta_1 {F^k}^T J_k (J_k^T J_k + \mu_k^* I)^{-1} J_k^T F^k \\
&\geq \frac{\eta_1}{2}\lambda_{\min}\Big((J_k^T J_k + \mu_k^* I)^{-1}\Big)\|J_k^T F^k\|^2 \\
&> \frac{\eta_1}{2(U_J^2 + U_\mu)}\varepsilon_g^2,
\end{aligned}$$

where the second equality follows from the definition of $d^k(\mu_k^*)$, and the last inequality follows from $\|J_k^T F^k\| > \varepsilon_g$ and (5.2.4). □

Now, we give an upper bound of K_{outer}.

Theorem 5.2.1. *Suppose that Assumptions* 5.2.1 *and* 5.2.2 *hold. Then,*

$$K_{\text{outer}} \leq \left\lceil \frac{f_0}{p}\varepsilon_g^{-2} + 1 \right\rceil.$$

Proof. Note that, $f(x) \geq 0$ from the definition of f. Then, we can prove the theorem in a way similar to the proof of Theorem 3.4.2. □

By using Theorem 5.2.1, we show the main theorem of the chapter.

Theorem 5.2.2. *Suppose that Assumptions* 5.2.1 *and* 5.2.2 *hold. Then,*

$$K_{\text{total}} \le \left\lceil \log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0 \gamma_1^{K_{\text{outer}}}}\right) + 1 \right\rceil,$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$.

Proof. Suppose the contrary, i.e., $K_{\text{total}} > \lceil \log_{\gamma_2}(U_\mu \gamma_2^{K_{\text{outer}}}/\mu_0 \gamma_1^{K_{\text{outer}}}) + 1 \rceil$. The number of satisfying $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) < \eta_1$ is $\sum_{k=0}^{K_{\text{outer}}-1}(l_k^* - 1)$. Moreover, the number of satisfying $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \ge \eta_2$ is at most K_{outer}. It then follows from the updating rule of μ_k that

$$\begin{aligned} \mu^*_{K_{\text{outer}}-1} &\ge \mu_0 \gamma_2^{\sum_{k=0}^{K_{\text{outer}}-1}(l_k^*-1)} \gamma_1^{K_{\text{outer}}} = \mu_0 \gamma_2^{K_{\text{total}}} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} \\ &> \mu_0 \gamma_2^{\log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0 \gamma_1^{K_{\text{outer}}}}\right)} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} = U_\mu, \end{aligned}$$

where the last inequality follows from the assumption that $K_{\text{total}} > \lceil \log_{\gamma_2}(U_\mu \gamma_2^{K_{\text{outer}}}/\mu_0 \gamma_1^{K_{\text{outer}}}) + 1 \rceil$. This contradicts Lemma 5.2.4. It then follows from Theorem 5.2.1 that

$$\begin{aligned} K_{\text{total}} &\le \left\lceil \log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0 \gamma_1^{K_{\text{outer}}}}\right) + 1 \right\rceil = \lceil K_{\text{outer}}(1 - \log_{\gamma_2}\gamma_1) + \log_{\gamma_2} U_\mu - \log_{\gamma_2}\mu_0 + 1 \rceil \\ &\le \left\lceil \left\lceil \frac{f_0}{p}\varepsilon_g^{-2} + 1 \right\rceil (1 - \log_{\gamma_2}\gamma_1) + \log_{\gamma_2} U_\mu - \log_{\gamma_2}\mu_0 + 1 \right\rceil, \end{aligned}$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$. □

5.3 Concluding remarks

In this chapter, we have shown that the global complexity bound of the LMM with the simple updating rule of μ_k is $O(\varepsilon_g^{-2})$. Note that, we do not assume any regularity of ∇F to obtain the bound.

There exist many updating rules of μ_k which accelerate a convergence of the LMM [45, 55, 68]. For example, we can replace the identity matrix I in $d^k(\mu_k)$ with a symmetric positive definite matrix D that takes some account of relative sizes of x_i. We can show that the global complexity bound of the modified LMM is also $O(\varepsilon_g^{-2})$. On the other hand, Dan, Yamashita and Fukushima [20] showed that the LMM with $\mu_k = \|F^k\|^\delta$ has superlinear convergence, where δ is a constant such that $\delta \ge 0$. In this chapter, we do not apply their approach for simplicity of the global complexity bound analysis. In the next chapter, we will apply it and investigate a global compelxity bound of the LMM using the generalized Jacobian.

Chapter 6

Global complexity bound analysis of the Levenberg-Marquardt method for nonsmooth equations and its application to the nonlinear complementarity problem

6.1 Introduction

We consider a system of nonsmooth equations

$$F(x) = 0, \tag{6.1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitz continuous mapping. When the system (6.1.1) has a solution, it is equivalent to the following nonlinear least squares problem.

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \ f(x) := \frac{1}{2}\|F(x)\|^2. \tag{6.1.2}$$

In this chapter, we assume that the least squares merit function f is continuously differentiable, though F is nonsmooth. The system (6.1.1) satisfying these assumptions includes important applications such as the nonlinear complementarity problem (NCP) and the Karush-Kuhn-Tacker (KKT) system [23]. For the system (6.1.1) or the NCP, the Levenberg-Marquardt method (LMM) is known to be an efficient solution method [22, 33, 36, 41, 42, 59, 69].

In Chapter 5, we investigated the global complexity bound of the LMM for the nonlinear least squares problem. Under the assumption that F is continuously differentiable, we showed that it is $O(\varepsilon_g^{-2})$ without any regularity assumption on F. However, we cannot directly apply this result to a system of nonsmooth equations.

In this chapter, we consider the generalized LMM for the nonsmooth equations that uses the generalized Jacobian of F. We show that it has the same bound $O(\varepsilon_g^{-2})$ as Chapter 5. Moreover, under some regularity assumption of the generalized Jacobian of F, we also show that the global complexity bound is $O(\log \varepsilon_f^{-1})$. By applying these results to the NCP, we get the global complexity bounds for the NCP. In particular, we can get a reasonable bound when the mapping involved in the NCP is a uniformly P-function.

This chapter is organized as follows. In the next section, we introduce the generalized LMM for nonsmooth equations. In Section 6.3, we give the global complexity bounds of the LMM. In Section 6.4, we apply the results on the bounds to the NCP. Finally, Section 6.5 concludes this chapter.

6.2 Generalized Levenberg-Marquardt method

In this section, we explain the generalized LMM for the system of nonsmooth equations (6.1.1). In what follows, we denote $J_k \in \partial F(x^k)$. Throughout this chapter, we need the following assumptions.

Assumption 6.2.1.

(a) *The vector mapping F is locally Lipschitz continuous.*

(b) *The least squares merit function f is continuously differentiable.*

As mentioned in Section 2.2.4, we can use the generalized Jacobian under Assumption 6.2.1 (a). Moreover, the system (6.1.1) satisfying Assumption 6.2.1 includes important applications such as the NCP and the KKT system.

For the k-th iterative point x^k, the generalized LMM adopts a search direction $d^k(\mu_k)$ defined by

$$d^k(\mu_k) := -(J_k^T J_k + \mu_k I)^{-1} J_k^T F^k,$$

where μ_k is a positive parameter. In Chapter 5, we adopt the Osborne's updating rule [55] of the regularization parameter μ_k for global convergence of the LMM. In this chapter, we also adopt his rule with the following little modification since it is known that the LMM with $\mu_k = \|F^k\|^\delta$ has rapid rate of convergence in a neighborhood of a solution [20]. We set μ_k as

$$\mu_k := \nu_k \|F^k\|^\delta,$$

and we control a positive parameter ν_k instead of μ_k. Here, δ is a given constant such that $\delta \geq 0$. Therefore, the generalized LMM with In what follows, we denote the search direction as $d^k(\nu_k)$ instead of $d^k(\mu_k)$.

We control ν_k as follows. Let $\phi_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a model function of f at x^k defined by

$$\phi_k(d, \nu) := \frac{1}{2}\|F^k + J_k d\|^2 + \frac{1}{2}\nu \|F^k\|^\delta \|d\|^2.$$

Let $\rho_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the ratio of the reduction of the merit function value to that of the model function value, i.e.,

$$\rho_k(d, \nu) := \frac{f_k - f(x^k + d)}{f_k - \phi_k(d, \nu)}.$$

If $\rho_k(d^k(\nu_k), \nu_k)$ is large, then we adopt $d^k(\nu_k)$ and decrease the parameter ν_k. On the other hand, if $\rho_k(d^k(\nu_k), \nu_k)$ is small, then we increase ν_k and compute $d^k(\nu_k)$ once again.

The precise description of the generalized LMM is as follows.

The Generalized Levenberg-Marquardt Method

Step 0 : Choose parameters $\varepsilon_g, \nu_0, \delta, \gamma_1, \gamma_2, \eta_1, \eta_2$ such that

$$0 < \varepsilon_g < 1,\ \nu_0 > 0,\ \delta \geq 0,\ \gamma_1 < 1 < \gamma_2,\ 0 < \eta_1 \leq \eta_2 \leq 1.$$

Choose a starting point x^0. Set $k := 0$.

Step 1 : Choose $J_k \in \partial F(x^k)$. If $\|J_k^T F^k\| \leq \varepsilon_g$, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\nu}_{l_k} := \nu_k$.

Step 2.1 : Compute $d^k(\bar{\nu}_{l_k})$.

Step 2.2 : Compute $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})$. If $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1$, then set $\bar{\nu}_{l_k+1} := \gamma_2 \bar{\nu}_{l_k}$ and $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $\eta_2 > \rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$, then set $\nu_{k+1} := \bar{\nu}_{l_k}$. If $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_2$, then set $\nu_{k+1} := \gamma_1 \bar{\nu}_{l_k}$. Set $x^{k+1} := x^k + d^k(\bar{\nu}_{l_k})$ and $k := k + 1$. Go to Step 1.

In what follows, for simplicity, we denote l_k and $\bar{\nu}_{l_k}$ at the last iteration of the inner loops of Steps 2.0–2.2 for each k as l_k^* and ν_k^*, respectively.

In the remainder of this section, we show that the generalized LMM is well-defined when $\|J_k^T F^k\| \neq 0$. First, we give a lower bound of the reduction of the model function.

Lemma 6.2.1. *Suppose that Assumption* 6.2.1 *holds. Then,*

$$f_k - \phi_k(d^k(\nu), \nu) = -\frac{1}{2} F^{k^T} J_k d^k(\nu) \geq \frac{\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu \|F^k\|^\delta)}.$$

Proof. By the definitions of f_k, $\phi_k(d^k(\nu), \nu)$ and $d^k(\nu)$, we have

$$\begin{aligned}
f_k - \phi_k(d^k(\nu), \nu) &= \frac{1}{2}\|F^k\|^2 - \left(\frac{1}{2}\|F^k + J_k d^k(\nu)\|^2 + \frac{1}{2}\nu \|F^k\|^\delta \|d^k(\nu)\|^2\right) \\
&= -F^{k^T} J_k d^k(\nu) - \frac{1}{2} d^k(\nu)^T (J_k^T J_k + \nu \|F^k\|^\delta I) d^k(\nu) \\
&= -\frac{1}{2} F^{k^T} J_k d^k(\nu) \\
&= \frac{1}{2} F^{k^T} J_k (J_k^T J_k + \nu \|F^k\|^\delta I)^{-1} J_k^T F^k \\
&\geq \frac{\lambda_{\min}\left((J_k^T J_k + \nu \|F^k\|^\delta I)^{-1}\right)}{2} \|J_k^T F^k\|^2 \\
&= \frac{\|J_k^T F^k\|^2}{2\lambda_{\max}(J_k^T J_k + \nu \|F^k\|^\delta I)} \\
&= \frac{\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu \|F^k\|^\delta)}.
\end{aligned}$$

This completes the proof. □

Next, we give an upper bound of $\|d^k(\nu)\|$.

Lemma 6.2.2. *Suppose that Assumption* 6.2.1 *holds. Then,*

$$\|d^k(\nu)\| \le \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta}.$$

Proof. By the definition of $d^k(\nu)$, we have

$$\begin{aligned}\|d^k(\nu)\| &= \|(J_k^T J_k + \nu\|F^k\|^\delta I)^{-1} J_k^T F^k\| \le \|(J_k^T J_k + \nu\|F^k\|^\delta I)^{-1}\| \cdot \|J_k^T F^k\| \\ &= \lambda_{\max}\left((J_k^T J_k + \nu\|F^k\|^\delta I)^{-1}\right)\|J_k^T F^k\| = \frac{1}{\lambda_{\min}(J_k^T J_k + \nu\|F^k\|^\delta I)}\|J_k^T F^k\| \\ &\le \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta},\end{aligned}$$

where the last inequality follows from the positive semidefiniteness of $J_k^T J_k$. □

From Lemmas 6.2.1 and 6.2.2, we give an upper bound of $f(x^k + d^k(\nu))$.

Lemma 6.2.3. *Suppose that Assumption 6.2.1 holds. Then,*

$$\begin{aligned}f(x^k + d^k(\nu)) \le \phi_k(d^k(\nu), \nu) &- \frac{\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu\|F^k\|^\delta)} \\ &+ \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta}\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau.\end{aligned}$$

Proof. Since f is continuously differentiable, we have

$$\begin{aligned}&f(x^k + d^k(\nu)) \\ &= f(x^k) + \int_0^1 \nabla f(x^k + \tau d^k(\nu))^T d^k(\nu)\mathrm{d}\tau \\ &= f(x^k) + \int_0^1 \nabla f(x^k + \tau d^k(\nu))^T d^k(\nu)\mathrm{d}\tau \\ &\quad + \phi_k(d^k(\nu), \nu) - \phi_k(d^k(\nu), \nu) + {F^k}^T J_k d^k(\nu) - {F^k}^T J_k d^k(\nu) \\ &= \phi_k(d^k(\nu), \nu) + (f(x^k) - \phi_k(d^k(\nu), \nu) + {F^k}^T J_k d^k(\nu)) \\ &\quad + \int_0^1 (\nabla f(x^k + \tau d^k(\nu)) - J_k^T F^k)^T d^k(\nu)\mathrm{d}\tau.\end{aligned}$$

It then follows from Lemma 6.2.1 and $g^k = J_k^T F^k$ that

$$\begin{aligned}&f(x^k + d^k(\nu)) \\ &= \phi_k(d^k(\nu), \nu) - (f(x^k) - \phi_k(d^k(\nu), \nu)) + \int_0^1 (\nabla f(x^k + \tau d^k(\nu)) - g^k)^T d^k(\nu)\mathrm{d}\tau \\ &\le \phi_k(d^k(\nu), \nu) - (f(x^k) - \phi_k(d^k(\nu), \nu)) + \|d^k(\nu)\|\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau \\ &\le \phi_k(d^k(\nu), \nu) - \frac{\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu\|F^k\|^\delta)} + \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta}\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau,\end{aligned}$$

where the last inequality follows from Lemmas 6.2.1 and 6.2.2. □

Next, we give the following key lemma for the well-definedness.

Lemma 6.2.4. *Suppose that Assumption* 6.2.1 *holds. Suppose also that* $\|J_k^T F^k\| \neq 0$. *Then,*

$$\rho_k(d^k(\nu), \nu) \geq 1$$

for ν *sufficiently large.*

Proof. Since $\|J_k^T F^k\| \neq 0$, we have $\|F^k\| \neq 0$. Thus, if ν is sufficiently large, $\nu\|F^k\|^\delta \geq \|J_k\|^2$ holds. In what follows, we suppose that $\nu\|F^k\|^\delta \geq \|J_k\|^2$ holds without loss of generality. It then follows from Lemma 6.2.3 that

$$\begin{aligned} f(x^k + d^k(\nu)) &\leq \phi_k(d^k(\nu), \nu) - \frac{\|J_k^T F^k\|^2}{4\nu\|F^k\|^\delta} + \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta} \int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\| \mathrm{d}\tau \\ &\leq \phi_k(d^k(\nu), \nu) + \frac{\|J_k^T F^k\|}{4\nu\|F^k\|^\delta} \left(-\|J_k^T F^k\| + 4 \int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\| \mathrm{d}\tau \right). \end{aligned} \tag{6.2.1}$$

Taking $\nu \to \infty$, we have $\lim_{\nu\to\infty} \|d^k(\nu)\| = 0$ from the definition of $d^k(\nu)$, and hence

$$\lim_{\nu\to\infty} \int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\| \mathrm{d}\tau = 0.$$

Thus, since $\|J_k^T F^k\| \neq 0$, the following inequality holds for sufficiently large ν.

$$4 \int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\| \mathrm{d}\tau \leq \|J_k^T F^k\|.$$

It then follows from (6.2.1) that

$$f(x^k + d^k(\nu)) \leq \phi_k(d^k(\nu), \nu).$$

Therefore, by the definition of $\rho_k(d^k(\nu), \nu)$, we have

$$\rho_k(d^k(\nu), \nu) = \frac{f(x^k) - f(x^k + d^k(\nu))}{f(x^k) - \phi_k(d^k(\nu), \nu)} \geq 1,$$

which is the desired inequality. □

Now, we show the well-definedness of the generalized LMM.

Theorem 6.2.1. *Suppose that Assumption* 6.2.1 *holds. Suppose also that* $\|J_k^T F^k\| \neq 0$. *Then, the generalized LMM is well-defined, i.e., the number* l_k *of inner iteration is finite.*

Proof. From the updating rule of $\bar{\nu}_{l_k}$, we have $\bar{\nu}_{l_k} \to \infty$ as $l_k \to \infty$. Thus, when l_k is sufficiently large, we have from Lemma 6.2.4 that

$$\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) = \frac{f_k - f(x^k + d^k(\bar{\nu}_{l_k}))}{f_k - \phi_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})} \geq 1 \geq \eta_1.$$

Therefore, the generalized LMM is well-defined. □

6.3 Global complexity bound

In this section, we estimate the global complexity bound of the generalized LMM. Let $K_{\rm outer}$ be the total number of outer iterations when the algorithm terminates. If there does not exist such $K_{\rm outer}$, we define $K_{\rm outer} := \infty$. Moreover, let $K_{\rm total}$ be the total number of inner iterations, until $k = K_{\rm outer}$ holds, i.e.,

$$K_{\rm total} := \sum_{k=0}^{K_{\rm outer}-1} l_k^*.$$

Note that, $K_{\rm total}$ means the total number of solving linear equations.

In order to investigate $K_{\rm total}$, we firstly make the following assumption.

Assumption 6.3.1.

(a) $\delta \le 1$.

(b) *The level set of f at the initial point x^0 is compact, i.e.,* $\Omega := \{x \in \mathbb{R}^n \mid f(x) \le f_0\}$ *is compact.*

Since $\{f_k\}$ is monotonically decreasing, the sequence $\{x^k\}$ is included in the compact set Ω. Moreover, since the generalized Jacobian ∂F is upper semi-continuous as mentioned in Section 2.2.4, there exist positive constants $F_{\max}$ and U_J such that

$$\|F(x)\| \le F_{\max},\ \|J\| \le U_J,\ \forall J \in \partial F(x),\ \forall x \in \Omega. \tag{6.3.1}$$

Now, we show that $\|d^k(\nu)\|$ is bounded from above when $\nu \in [\nu_0, \infty)$.

Lemma 6.3.1. *Suppose that Assumptions* 6.2.1 *and* 6.3.1 *hold. Then, for any* $\nu \in [\nu_0, \infty)$,

$$\|d^k(\nu)\| \le U_d,$$

where

$$U_d := \frac{U_J F_{\max}^{1-\delta}}{\nu_0}.$$

Proof. It follows from Lemma 6.2.2 that

$$\|d^k(\nu)\| \le \frac{\|J_k^T F^k\|}{\nu \|F^k\|^\delta} \le \frac{\|J_k^T\| \cdot \|F^k\|}{\nu \|F^k\|^\delta} \le \frac{U_J F_{\max}^{1-\delta}}{\nu_0},$$

where the last inequality follows from (6.3.1) and $\nu \ge \nu_0$. □

In Chapter 5, when F is continuously differentiable, we assumed that the Jacobian of F is Lipschitz continuous to investigate the global complexity bound of the LMM. However, since F is nonsmooth in this chapter, the assumption does not hold in general. Instead, we assume that the gradient of the merit function f is Lipschitz continuous.

Assumption 6.3.2. *Let* $U_d := U_J F_{\max}^{1-\delta}/\nu_0$. ∇f *is Lipschitz continuous on* $\Omega + B(0, U_d)$ *with modulus* L_g.

By using the assumption, we show that $\rho_k(d^k(\nu),\nu) \geq 1$ if ν is greater than a specific value depending on F^k.

Lemma 6.3.2. *Suppose that Assumptions* 6.2.1, 6.3.1 *and* 6.3.2 *hold. Suppose also that*

$$\nu \geq \frac{\max(U_J^2, \nu_0 F_{\max}^\delta, 2L)}{\|F^k\|^\delta}.$$

Then,

$$\rho_k(d^k(\nu),\nu) \geq 1.$$

Proof. From (6.3.1) and the assumption on ν, we have the following three inequalities.

$$\nu\|F^k\|^\delta \geq U_J^2 \geq \|J_k\|^2, \tag{6.3.2}$$

$$\nu \geq \frac{\nu_0 F_{\max}^\delta}{\|F^k\|^\delta} \geq \nu_0, \tag{6.3.3}$$

$$\nu\|F^k\|^\delta \geq 2L. \tag{6.3.4}$$

By using (6.3.2) and Lemma 6.2.3, we have

$$\begin{aligned} f(x^k + d^k(\nu)) &\leq \phi_k(d^k(\nu),\nu) - \frac{\|J_k^T F^k\|^2}{4\nu\|F^k\|^\delta} + \frac{\|J_k^T F^k\|}{\nu\|F^k\|^\delta}\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau \\ &\leq \phi_k(d^k(\nu),\nu) + \frac{\|J_k^T F^k\|}{4\nu\|F^k\|^\delta}\left(-\|J_k^T F^k\| + 4\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau\right). \end{aligned} \tag{6.3.5}$$

On the other hand, by using (6.3.3) and Lemma 6.3.1, we have $x^k + \tau d^k(\nu) \in \Omega + B(0, U_d)$ for any $\tau \in [0,1]$. It then follows from Assumption 6.3.2 that

$$\begin{aligned} 4\int_0^1 \|(\nabla f(x^k + \tau d^k(\nu)) - g^k)\|\mathrm{d}\tau &\leq 4L\|d^k(\nu)\|\int_0^1 \tau\mathrm{d}\tau = 2L\|d^k(\nu)\| \\ &\leq \frac{2L\|J_k^T F^k\|}{\nu\|F^k\|^\delta} \leq \|J_k^T F^k\|, \end{aligned}$$

where the second inequality follows from Lemma 6.2.2, and the last inequality follows from (6.3.4). It then follows from (6.3.5) that

$$f(x^k + d^k(\nu)) \leq \phi_k(d^k(\nu),\nu).$$

Therefore, by the definition of $\rho_k(d^k(\nu),\nu)$, we have

$$\rho_k(d^k(\nu),\nu) = \frac{f_k - f(x^k + d^k(\nu))}{f_k - \phi_k(d^k(\nu),\nu)} \geq 1,$$

which is the desired inequality. □

From Lemma 6.3.2, we can show that $\nu_k^*\|F^k\|^\delta$ is bounded from above.

Lemma 6.3.3. *Suppose that Assumptions* 6.2.1, 6.3.1 *and* 6.3.2 *hold. Then,*

$$\nu_k^*\|F^k\|^\delta \le U_{\nu F},$$

where

$$U_{\nu F} := \gamma_2 \max(U_J^2, \nu_0 F_{\max}^\delta, 2L).$$

Proof. From Lemma 6.3.2, if $\bar{\nu}_{l_k}\|F^k\|^\delta \ge \max(U_J^2, \nu_0 F_{\max}^\delta, 2L)$, then $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \ge 1$, and hence the inner loops of Step 2 must terminate. Therefore, if $\bar{\nu}_1\|F^k\|^\delta \ge \max(U_J^2, \nu_0 F_{\max}^\delta, 2L)$ at the k-th iteration, then $\nu_k^*\|F^k\|^\delta = \bar{\nu}_1\|F^k\|^\delta$. On the other hand, if $\bar{\nu}_1\|F^k\|^\delta < \max(U_J^2, \nu_0 F_{\max}^\delta, 2L)$, then $\nu_k^*\|F^k\|^\delta$ must satisfy $\nu_k^*\|F^k\|^\delta \le \gamma_2 \max(\nu_0 F_{\max}^\delta, 2L)$. Otherwise, $\bar{\nu}_{l_k^*-1}\|F^k\|^\delta > \max(U_J^2, \nu_0 F_{\max}^\delta, 2L)$, which contradicts $\rho_k(d^k(\bar{\nu}_{l_k^*-1}), \bar{\nu}_{l_k^*-1}) < \eta_1 \le 1$. Consequently, we have

$$\begin{aligned}
\nu_k^*\|F^k\|^\delta &\le \max(\bar{\nu}_1\|F^k\|^\delta, \gamma_2 U_J^2, \gamma_2\nu_0 F_{\max}^\delta, \gamma_2 2L) \\
&= \max(\nu_{k-1}^*\|F^k\|^\delta, \gamma_2 U_J^2, \gamma_2\nu_0 F_{\max}^\delta, \gamma_2 2L) \\
&\le \max(\nu_{k-1}^*\|F^{k-1}\|^\delta, \gamma_2 U_J^2, \gamma_2\nu_0 F_{\max}^\delta, \gamma_2 2L) \\
&\le \cdots \le \max(\nu_0\|F^0\|^\delta, \gamma_2 U_J^2, \gamma_2\nu_0 F_{\max}^\delta, \gamma_2 2L) \\
&= \gamma_2 \max(\gamma_2 U_J^2, \nu_0 F_{\max}^\delta, 2L)
\end{aligned}$$

from the updating rule of ν. □

By using the above lemma, we give a lower bound of the reduction of the merit function when $k < K_{\text{outer}}$.

Lemma 6.3.4. *Suppose that Assumptions* 6.2.1, 6.3.1 *and* 6.3.2 *hold. Suppose also that* $k < K_{\text{outer}}$. *Then,*

$$f_k - f_{k+1} > p\varepsilon_g^2.$$

where

$$p := \frac{\eta_1}{2(U_J^2 + U_{\nu F})}.$$

Proof. Since $\rho_k(d^k(\nu_k^*), \nu_k^*) \ge \eta_1$ from the definition of ν_k^*, we have

$$f_k - f_{k+1} \ge \eta_1(f_k - \phi_k(d^k(\nu_k^*), \nu_k^*)) \ge \frac{\eta_1\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu_k^*\|F^k\|^\delta)}, \tag{6.3.6}$$

where the last inequality follows from Lemma 6.2.1. On the other hand, we have $\|J_k^T F^k\| > \varepsilon_g$, $\forall k < K_{\text{outer}}$ from the definition of K_{outer}. It then follows from Lemma 6.3.3, (6.3.1) and (6.3.6) that

$$f_k - f_{k+1} \ge \frac{\eta_1\|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu_k^*\|F^k\|^\delta)} \ge \frac{\eta_1}{2(U_J^2 + U_{\nu F})}\varepsilon_g^2,$$

which is the desired inequality. □

Now, we give an upper bound of K_{outer}.

Theorem 6.3.1. *Suppose that Assumptions* 6.2.1, 6.3.1 *and* 6.3.2 *hold. Then,*

$$K_{\text{outer}} \le \left\lceil \frac{f_0}{p}\varepsilon_g^{-2} + 1 \right\rceil.$$

Proof. Note that, $f(x) \ge 0$ from the definition of f. Then, we can prove the theorem in a way similar to the proof of Theorem 3.4.2. □

From Theorem 6.3.1, the next theorem gives the global complexity bound K_{total} of the LMM.

Theorem 6.3.2. *Suppose that Assumptions 6.2.1, 6.3.1 and 6.3.2 hold. Then,*

$$K_{\text{total}} \le \left\lceil \log_{\gamma_2}\left(\frac{U_{\nu F} U_J^{\delta} \gamma_2^{K_{\text{outer}}}}{\nu_0 \gamma_1^{K_{\text{outer}}}} \varepsilon_g^{-\delta} \right) + 1 \right\rceil,$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$.

Proof. Since $\varepsilon_g < \|J_{K_{\text{outer}}-1}^T F^{K_{\text{outer}}-1}\| \le U_J \|F^{K_{\text{outer}}-1}\|$ from (6.3.1), we have

$$\|F^{K_{\text{outer}}-1}\| > \frac{\varepsilon_g}{U_J}.$$

Now we suppose the contrary of the theorem, i.e.,

$$K_{\text{total}} > \lceil \log_{\gamma_2}(\varepsilon_g^{-\delta} U_{\nu F} U_J^{\delta} \gamma_2^{K_{\text{outer}}} / \nu_0 \gamma_1^{K_{\text{outer}}}) + 1 \rceil. \tag{6.3.7}$$

The number of satisfying $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1$ is $\sum_{k=0}^{K_{\text{outer}}-1}(l_k^* - 1)$. Moreover, the number of satisfying $\rho_k(d^k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \ge \eta_2$ is at most K_{outer}. It then follows from the updating rule of ν_k that

$$\begin{aligned}
\nu_{K_{\text{outer}}-1}^* \|F^{K_{\text{outer}}-1}\|^{\delta} &> \nu_{K_{\text{outer}}-1}^* U_J^{-\delta} \varepsilon_g^{\delta} \\
&= \nu_0 \gamma_2^{\sum_{k=0}^{K_{\text{outer}}-1}(l_k^*-1)} \gamma_1^{K_{\text{outer}}} U_J^{-\delta} \varepsilon_g^{\delta} = \nu_0 \gamma_2^{K_{\text{total}}} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} U_J^{-\delta} \varepsilon_g^{\delta} \\
&> \nu_0 \gamma_2^{\log_{\gamma_2}\left(\frac{U_{\nu F} U_J^{\delta} \gamma_2^{K_{\text{outer}}}}{\nu_0 \gamma_1^{K_{\text{outer}}}} \varepsilon_g^{-\delta} \right)} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} \varepsilon_g^{\delta} = U_{\nu F},
\end{aligned}$$

where the last inequality follows from the assumption (6.3.7). This contradicts Lemma 6.3.3. It then follows from Theorem 6.3.1 that

$$\begin{aligned}
K_{\text{total}} &\le \left\lceil \log_{\gamma_2}\left(\frac{U_{\nu F} U_J^{\delta} \gamma_2^{K_{\text{outer}}}}{\nu_0 \gamma_1^{K_{\text{outer}}}} \varepsilon_g^{-\delta} \right) + 1 \right\rceil \\
&= \lceil K_{\text{outer}}(1 - \log_{\gamma_2} \gamma_1) + \log_{\gamma_2} U_{\nu F} + \delta \log_{\gamma_2} U_J + \delta \log_{\gamma_2} \varepsilon_g^{-1} - \log_{\gamma_2} \nu_0 + 1 \rceil \\
&\le \left\lceil \left\lceil \frac{f_0}{p}\varepsilon_g^{-2} + 1 \right\rceil (1 - \log_{\gamma_2} \gamma_1) + \log_{\gamma_2} U_{\nu F} + \delta \log_{\gamma_2} U_J + \delta \log_{\gamma_2} \varepsilon_g^{-1} - \log_{\gamma_2} \nu_0 + 1 \right\rceil,
\end{aligned}$$

and hence $K_{\text{total}} = O(\varepsilon_g^{-2})$. □

Note that, since $J_k^T F^k = 0$ does not imply $F^k = 0$, Theorem 6.3.2 does not provide the global complexity bound of $f_k = \frac{1}{2}\|F^k\|^2 \le \varepsilon_f$ for some positive constant ε_f. To get the bound, we replace the termination criterion in Step 1 with $f_k \le \varepsilon_f$ in the remainder of this section. We call the resulting method the modified LMM, and denote the total number of inner iterations of the modified LMM as $\widehat{K}_{\text{total}}$. Note that, since f is nonconvex, the modified LMM may not terminate. Thus, we further assume a regularity of the generalized Jacobian.

Assumption 6.3.3. *There exists a positive constant* σ *such that* $\lambda_{\min}(J_k J_k^T) \ge \sigma$ *for all* $k \ge 0$.

Under Assumption 6.3.3, we give the global complexity bound $\widehat{K}_{\text{total}}$.

Theorem 6.3.3. *Suppose that Assumptions* 6.2.1*,* 6.3.1*,* 6.3.2 *and* 6.3.3 *hold. Then,* $\widehat{K}_{\text{total}} = O(\log \varepsilon_f^{-1})$.

Proof. Since $\rho_k(d^k(\nu_k^*), \nu_k^*) \ge \eta_1$ from the definition of ν_k^*,

$$f_k - f_{k+1} \ge \eta_1(f_k - \phi_k(d^k(\nu_k^*), \nu_k^*)) \ge \frac{\eta_1 \|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu_k^* \|F^k\|^\delta)}, \tag{6.3.8}$$

where the last inequality follows from Lemma 6.2.1. On the other hand, Assumption 6.3.3 implies that $\|J_k^T F^k\|^2 \ge \sigma \|F^k\|^2$. It then follows from (6.3.8) that

$$\begin{aligned} f_k - f_{k+1} &\ge \frac{\eta_1 \|J_k^T F^k\|^2}{2(\|J_k\|^2 + \nu_k^* \|F^k\|^\delta)} \ge \frac{\eta_1 \sigma}{2(\|J_k\|^2 + \nu_k^* \|F^k\|^\delta)} \|F^k\|^2 \\ &\ge \frac{\eta_1 \sigma}{2(U_J^2 + U_{\nu F})} \|F^k\|^2 = \frac{\eta_1 \sigma}{U_J^2 + U_{\nu F}} f_k. \end{aligned}$$

where the third inequality follows from (6.3.1) and Lemma 6.3.3, and the last equality follows from the definition of f_k. Therefore, we have

$$f_k \le \left(1 - \frac{\eta_1 \sigma}{U_J^2 + U_{\nu F}}\right) f_{k-1} \le \left(1 - \frac{\eta_1 \sigma}{U_J^2 + U_{\nu F}}\right)^k f_0,$$

and hence if

$$k \ge \frac{\log \frac{f_0}{\varepsilon_f}}{\log \left(1 - \frac{\eta_1 \sigma}{U_J^2 + U_{\nu F}}\right)^{-1}}$$

then $f_k \le \varepsilon_f^2$. Thus, we have $\widehat{K}_{\text{outer}} = O(\log \varepsilon_f^{-1})$, where $\widehat{K}_{\text{outer}}$ is the total number of outer iterations of the modified LMM. It then directly follows from Theorem 6.3.2 that $\widehat{K}_{\text{total}} = O(\log \varepsilon_f^{-1})$. □

6.4 Application to the nonlinear complementarity problem

We apply the results obtained in the previous section to the NCP(G) [23]: Find $x \in \mathbb{R}^n$ such that

$$x \ge 0,\ G(x) \ge 0,\ x^T G(x) = 0,$$

where $G : \mathbb{R}^n \to \mathbb{R}^n$. In this section, we assume that the mapping G satisfy the following assumptions.

Assumption 6.4.1.

(a) *The vector mapping G is continuously differentiable.*

(b) *∇G is locally Lipschitz continuous.*

By using the Fischer-Burmeister function, we can reformulate NCP(G) into the following non-smooth equations [26].

$$F(x) := \begin{pmatrix} \psi(x_1, G_1(x)) \\ \vdots \\ \psi(x_n, G_n(x)) \end{pmatrix} = 0, \tag{6.4.1}$$

where $\psi : \mathbb{R}^2 \to \mathbb{R}$ is the Fischer-Burmeister function defined by

$$\psi(a, b) := \sqrt{a^2 + b^2} - a - b.$$

Note that, ψ is not differentiable at $(0, 0)$. Therefore, if there exists i such that $x_i = G_i(x) = 0$, then F is not differentiable at x. Nevertheless, F is locally Lipschitz continuous under Assumption 6.4.1 [24]. Moreover, the least squares merit function $f(x) := \frac{1}{2}\|F(x)\|^2$ has the following properties [14].

Lemma 6.4.1. *Suppose that Assumption* 6.4.1 *holds.*

(a) *F is locally Lipschitz continuous.*

(b) *f is continuously differentiable.*

(c) *∇f is locally Lipschitz continuous.*

Lemma 6.4.1 (c) implies that ∇f is Lipschitz continuous on any compact set.

By using Lemma 6.4.1, we get the global complexity bound of the generalized LMM for the equations (6.4.1) equivalent to the NCP as a direct application of Theorem 6.3.2.

Theorem 6.4.1. *Suppose that Assumption* 6.4.1 *holds. Suppose also that $\delta \leq 1$ and a sequence generated by the generalized LMM is bounded. Then, the global complexity bound of the generalized LMM for the NCP is $O(\varepsilon_g^{-2})$.*

Remark 6.4.1. *A sequence generated by the generalized LMM is bounded if the level set of f is compact. The level set of f is compact if G is a uniformly P-function* [24] (*see Assumption* 6.4.2 *for the definition*). *The level set of f is also compact if G is monotone, NCP(G) has a strictly feasible solution and the Fischer-Burmeister function is replaced with the penalized Fischer-Burmeister function $\psi_\tau(a, b) = \tau\psi(a, b) + (1 - \tau)\max(0, a)\max(0, b)$, where $\tau \in (0, 1)$ is an arbitrary but fixed constant* [13].

Remark 6.4.2. *Note that, the bound in Theorem* 6.4.1 *is not for a solution of NCP(G) but for a stationary point of f. However, a stationary point of f is a solution of NCP(G) if G is P_0-function, i.e., there exists i such that $x_i \neq y_i$ and $(x_i - y_i)(G_i(x) - G_i(y)) \geq 0, \forall x, y \in \mathbb{R}^n$* [24].

Next, as related to Assumption 6.3.3, we further make the following assumption on G.

Assumption 6.4.2. *G is a uniformly P-function, i.e., there exists a positive constant $\alpha > 0$ such that*

$$\max_{1\le i\le n}(x_i - y_i)(G_i(x) - G_i(y)) \ge \alpha\|x-y\|^2, \forall x, y \in \mathbb{R}^n.$$

When G is a uniformly P-function, it is well-known that the following properties hold [24, 32, 35, 44].

Lemma 6.4.2. *Suppose that Assumptions 6.4.1 and 6.4.2 hold.*

(a) *The level set Ω of the merit function f is compact.*

(b) *For any $J \in \partial F(x)$ and $x \in \mathbb{R}^n$, J is nonsingular.*

(c) *The NCP(G) has a unique solution x^*.*

(d) *There exists a positive constant κ such that $\|x - x^*\| \le \kappa\|F(x)\|$ for any $x \in \Omega$.*

From Lemma 6.4.2 (a), (b) and the upper semi-continuity of the generalized Jacobian ∂F, there exists a positive constant σ such that $\lambda_{\min}(JJ^T) \ge \sigma$ for any $J \in \partial F(x)$ and $x \in \Omega$. Therefore, Assumption 6.3.3 holds.

Now, we get the bound for the NCP(G) as a direct application of Theorem 6.3.3.

Theorem 6.4.2. *Suppose that Assumptions 6.4.1 and 6.4.2 hold. Suppose also that $\delta \le 1$. Then, the global complexity bound of the modified LMM defined in Section 6.3 is $O(\log \varepsilon_f^{-1})$. Moreover, for an approximate solution $\hat{x}$ such that $f(\hat{x}) \le \varepsilon_f$, the distance $\|\hat{x} - x^*\| = O(\varepsilon_f^{\frac{1}{2}})$.*

Proof. The first part of the theorem directly follows from Theorem 6.3.3. The second part of the theorem follows from Lemma 6.4.2 (d) and the assumption on $\hat{x}$. □

6.5 Concluding remarks

In this chapter, we have investigated the global complexity bound of the generalized LMM for the nonsmooth equations. We have shown that the bound is $O(\varepsilon_g^{-2})$ without any regularity or convex assumptions. We have also shown that the bound is $O(\log \varepsilon_f^{-1})$ under the regularity assumption of the generalized Jacobian. Moreover, by applying these results to the NCP, we have obtained the same global complexity bounds of the generalized LMM for the NCP.

In this chapter, we have assumed that the mapping G involved in the NCP is a uniformly P-function for the regularity assumption of the generalized Jacobian. By using other assumption such as the monotonicity of G, we may have a better global complexity bound.

Chapter 7

Conclusion

In this thesis, we have studied on the regularized Newton-type methods for unconstrained minimization problems and their complexities. The results obtained in this thesis are summarized as follows.

- In Chapter 3, we have extended the regularized Newton method (RNM) to the unconstrained nonconvex minimization problem. To guarantee global convergence, we have exploited Armijo's step size rule. We have shown that the extended RNM has global convergence, and superlinear convergence under the local error bound condition. We have also shown that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumptions that the level set of f at the initial point x^0 is compact and $\nabla^2 f$ is Lipschitz continuous on the set.

- In Chapter 4, we have proposed another type of RNM for the unconstrained nonconvex minimization problem. The proposed RNM (called the adaptive RNM) does not use any line search, but controls the regularization parameter for global convergence. We have shown that the adaptive RNM has global convergence, and superlinear convergence under the local error bound condition. Moreover, under the assumptions that the level set of f at the initial point x^0 is compact and $\nabla^2 f$ is Lipschitz continuous on the set, we have shown that the global complexity bounds are $O(\varepsilon_g^{-2})$ when f is nonconvex, $O(\varepsilon_g^{-\frac{5}{3}})$ and $O(\varepsilon_f^{-\frac{2}{3}})$ when f is convex, and $O(\varepsilon_g^{-1})$ and $O(\log \varepsilon_f^{-1})$ when f is strongly convex. In addition, when f is not convex but quasi-convex and analytic, we have given the better global complexity bound than $O(\varepsilon_g^{-2})$. Finally, we have reported effectiveness of the adaptive RNM by some numerical experiments.

- In Chapter 5, we have investigated the global complexity bound of the Levenberg-Marquardt method (LMM) for the nonlinear least squares problem (1.1.4). The method uses Osborne's updating rule [55] which controls the regularization parameter μ_k directly for global convergence. We have shown that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumptions that the level set of the least squares merit function f at the initial point x^0 is compact and ∇F is Lipschitz continuous on the set.

- In Chapter 6, we have investigated the global complexity bound of the generalized LMM for the system of nonsmooth equations. The method also exploits Osborne's updating rule

[55] for global convergence. We have shown that the global complexity bound is $O(\varepsilon_g^{-2})$ under the assumptions that the level set of the least squares merit function f at the initial point x^0 is compact and ∇f is Lipschitz continuous on the set. In addition, by assuming the regularity of the generalized Jacobian, we have also shown that the global complexity bound is $O(\log \varepsilon_f^{-1})$. Furthermore, we have applied these results to nonsmooth equations equivalent to the nonlinear complementarity problem (NCP), and analyzed the global complexity bounds for the NCP. In particular, we have given the reasonable bound when the mapping involved in the NCP is a uniformly P-function.

The results for the global complexity bound including our results are summarized as the following table.

Table 7.1: The results for the global complexity bound

Method	Bound (nonconvex)	Bound (convex)	Lipschitz
Steepest descent method	$O(\varepsilon_g^{-2})$	$O(\varepsilon_f^{-1})$	∇f
Accelerated steepest descent method	–	$O(\varepsilon_f^{-\frac{1}{2}})$	∇f
Trust-region NM	$O(\varepsilon_g^{-2})$	–	$\nabla^2 f$
(Polyak's) RNM	–	$O(\varepsilon_g^{-4})$	∇f
Cubic RNM	$O(\varepsilon_g^{-\frac{3}{2}})$	$O(\varepsilon_f^{-\frac{1}{2}})$	$\nabla^2 f$
Accelerated cubic RNM	–	$O(\varepsilon_f^{-\frac{1}{3}})$	$\nabla^2 f$
Extended RNM	$O(\varepsilon_g^{-2})$	–	$\nabla^2 f$
Adaptive RNM	$O(\varepsilon_g^{-2})$	$O(\varepsilon_g^{-\frac{5}{3}})$, $O(\varepsilon_f^{-\frac{2}{3}})$	$\nabla^2 f$
LMM	$O(\varepsilon_g^{-2})$	–	∇F
Generalized LMM	$O(\varepsilon_g^{-2})$	–	∇f

As we summarized above, we have made several contributions to the studies on the regularized Newton-type methods for unconstrained minimization problems and their global complexity bounds. However, there are some problems that remain unsolved. In the following, we list some future works.

- Under appropriate conditions, the RNM and the LMM have global convergence to a stationary point. When x^k is a stationary point, $d^k(\mu_k) = 0$ by the definition of $d^k(\mu_k)$. Thus, when f is nonconvex, they may not have global convergence to a point x^* that satisfies the second-order necessary optimality condition, i.e., $\nabla^2 f(x^*)$ is positive semidefinite. On the other hand, the trust-region NM and the cubic RNM have global convergence to such point [17, 53]. Therefore, it would be important to develop an RNM and an LMM that have global convergence to not only a stationary point, but also a point satisfying the second-order necessary optimality condition.
- Recently, by using the polynomial Hermite interpolation, Cartis, Gould and Toint [9] constructed problems where the global complexity bounds of some solution methods are tight.

Thus, by using their technique, we may be able to show the tightness of the global complexity bounds for the regularized Newton-type methods.

- As described in Section 1.4, the steepest descent method and the cubic RNM can be accelerated when f is convex [49, 50, 52]. By using the acceleration techniques in [49, 50, 52], we may be able to accelerate the RNM and give a better global complexity bound than $O(\varepsilon_f^{-\frac{2}{3}})$, which is obtained in Section 4.4.

- We proposed two RNMs for the unconstrained nonconvex minimization problem. It is challenging to study an RNM for the general nonconvex optimization problem that has both equality constraints and inequality constraints, and to show its good convergence properties. In particular, for the constrained optimization problem, the worst computational complexity of iterative methods has been discussed [3, 4, 7, 11, 48, 50, 52, 61]. These existing techniques may help us to extend our results on the global complexity bounds to the constrained optimization problem.

Appendix A

Table in Subsection 4.6.2

Table A.1: Comparison among solution methods

		Adaptive RNM		Extended RNM		
Name	n	N_f, N_L	f	N_f	N_L	f
3PK	30	6	1.72E + 00	333	333	1.72E + 00
AKIVA	2	6	6.17E + 00	7	7	6.17E + 00
ALLINITU	4	8	5.74E + 00	11	9	5.74E + 00
ARGLINA	200	4	2.00E + 02	8	8	2.00E + 02
ARWHEAD	100	5	8.79E − 14	6	6	0.00E + 00
BARD	3	7	8.21E − 03	10	10	8.21E − 03
BDQRTIC	100	9	3.79E + 02	10	10	3.79E + 02
BEALE	2	8	8.23E − 13	8	8	2.50E − 11
BIGGS6	6	100	8.17E − 08	102	97	1.11E − 06
BOX3	3	7	2.28E − 12	9	9	4.05E − 11
BRKMCC	2	3	1.69E − 01	4	4	1.69E − 01
BROWNAL	200	4	3.80E − 13	4	4	3.21E − 09
BROWNBS	2	12	1.00E − 13	22	20	0.00E + 00
BROWNDEN	4	8	8.58E + 04	8	8	8.58E + 04
BROYDN7D	100	33	3.28E + 01	30	30	3.28E + 01
BRYBND	100	12	3.56E − 15	11	9	3.61E − 13
CHNROSNB	50	119	3.55E − 19	51	43	1.79E − 17
CLIFF	2	27	2.00E − 01	34	34	2.00E − 01
COSINE	100	9	−9.90E + 01	12	11	−9.90E + 01
CRAGGLVY	100	13	3.23E + 01	15	15	3.23E + 01
CUBE	2	46	6.77E − 15	34	29	4.85E − 13
CURLY10	100	26	−1.00E + 04	22	20	−1.00E + 04
CURLY20	100	24	−1.00E + 04	20	18	−1.00E + 04
DECONVU	61	25	7.31E − 08	17	15	1.46E − 07
DENSCHNA	2	5	1.35E − 12	6	6	1.02E − 15

Table A.1: Comparison among solution methods

		Adaptive RNM		Extended RNM		
Name	n	N_f, N_L	f	N_f	N_L	f
DENSCHNB	2	5	1.85E − 20	4	4	4.77E − 18
DENSCHNC	2	10	4.87E − 20	11	11	2.27E − 17
DENSCHND	3	36	1.82E − 08	29	29	1.35E − 08
DENSCHNE	3	13	7.78E − 13	26	26	1.01E − 18
DENSCHNF	2	6	6.86E − 22	6	6	8.44E − 20
DIXMAANA	300	9	1.00E + 00	10	10	1.00E + 00
DIXMAANB	300	17	1.00E + 00	11	11	1.00E + 00
DIXMAANC	300	9	1.00E + 00	12	12	1.00E + 00
DIXMAAND	300	9	1.00E + 00	12	12	1.00E + 00
DIXMAANE	300	8	1.00E + 00	25	25	1.00E + 00
DIXMAANF	300	12	1.00E + 00	21	21	1.00E + 00
DIXMAANG	300	13	1.00E + 00	21	21	1.00E + 00
DIXMAANH	300	14	1.00E + 00	22	22	1.00E + 00
DIXMAANI	300	12	1.00E + 00	41	41	1.00E + 00
DIXMAANJ	300	21	1.00E + 00	27	27	1.00E + 00
DIXMAANK	15	18	1.00E + 00	17	17	1.00E + 00
DIXMAANL	300	22	1.00E + 00	27	27	1.00E + 00
DIXON3DQ	100	5	4.96E − 14	35	35	2.39E − 08
DQDRTIC	100	4	2.67E − 25	6	6	2.99E − 18
EDENSCH	36	12	2.19E + 02	12	12	2.19E + 02
ENGVAL1	100	7	1.09E + 02	8	8	1.09E + 02
ENGVAL2	3	17	1.28E − 14	21	21	8.39E − 22
ERRINROS	50	57	3.99E + 01	130	127	3.99E + 01
EXPFIT	2	12	2.41E − 01	10	8	2.41E − 01
FLETCBV2	100	3	−5.14E − 01	4	4	−5.14E − 01
FREUROTH	100	13	1.20E + 04	15	12	1.20E + 04
GENROSE	100	144	1.00E + 00	105	77	1.00E + 00
GROWTHLS	3	184	1.00E + 00	366	366	1.00E + 00
GULF	3	36	2.84E − 11	119	117	2.41E − 06
HAIRY	2	70	2.00E + 01	78	60	2.00E + 01
HATFLDD	3	21	6.62E − 08	21	21	6.76E − 08
HATFLDE	3	17	5.12E − 07	23	23	5.12E − 07
HEART6LS	6	1875	7.93E − 23	3193	2923	1.75E − 10
HEART8LS	8	175	4.00E − 23	107	83	1.90E − 19
HELIX	3	10	3.74E − 23	11	11	1.91E − 13
HIELOW	3	10	8.74E + 02	7	6	8.74E + 02

Table A.1: Comparison among solution methods

		Adaptive RNM		Extended RNM		
Name	n	N_f, N_L	f	N_f	N_L	f
HILBERTA	2	4	$3.39\text{E}-15$	9	9	$1.92\text{E}-13$
HILBERTB	10	3	$1.23\text{E}-12$	5	5	$2.44\text{E}-19$
HIMMELBB	2	12	$1.99\text{E}-18$	12	12	$6.13\text{E}-26$
HIMMELBF	4	158	$3.19\text{E}+02$	993	993	$3.19\text{E}+02$
HIMMELBG	2	7	$1.05\text{E}-14$	6	6	$1.17\text{E}-12$
HIMMELBH	2	6	$-1.00\text{E}+00$	7	6	$-1.00\text{E}+00$
HUMPS	2	340	$3.39\text{E}-12$	1275	1221	$4.20\text{E}-13$
KOWOSB	4	12	$3.08\text{E}-04$	8	8	$3.08\text{E}-04$
LIARWHD	100	10	$1.52\text{E}-13$	12	12	$1.19\text{E}-12$
LOGHAIRY	2	51	$6.53\text{E}+00$	214	211	$6.48\text{E}+00$
MARATOSB	2	–	–	948	672	$-1.00\text{E}+00$
MEXHAT	2	44	$-4.00\text{E}-02$	31	28	$-4.00\text{E}-02$
MOREBV	100	2	$7.69\text{E}-07$	3	3	$5.44\text{E}-07$
NONCVXU2	100	36	$2.32\text{E}+02$	98	98	$2.34\text{E}+02$
NONCVXUN	100	27	$2.34\text{E}+02$	42	42	$2.37\text{E}+02$
NONDIA	100	10	$4.93\text{E}-16$	8	7	$6.78\text{E}-21$
OSBORNEA	5	59	$5.51\text{E}-05$	52	32	$5.51\text{E}-05$
OSBORNEB	11	17	$4.01\text{E}-02$	26	26	$4.01\text{E}-02$
PALMER1C	8	6	$9.76\text{E}-02$	282	282	$9.76\text{E}-02$
PALMER1D	7	6	$6.53\text{E}-01$	189	189	$6.53\text{E}-01$
PALMER2C	8	6	$1.44\text{E}-02$	165	165	$1.44\text{E}-02$
PALMER3C	8	6	$1.95\text{E}-02$	170	170	$1.95\text{E}-02$
PALMER4C	8	6	$5.03\text{E}-02$	227	227	$5.03\text{E}-02$
PALMER5C	6	4	$2.13\text{E}+00$	7	7	$2.13\text{E}+00$
PALMER6C	8	6	$1.64\text{E}-02$	365	365	$1.64\text{E}-02$
PALMER7C	8	6	$6.02\text{E}-01$	1139	1139	$6.02\text{E}-01$
PALMER8C	8	6	$1.60\text{E}-01$	495	495	$1.60\text{E}-01$
PFIT1LS	3	613	$1.14\text{E}-10$	808	357	$1.85\text{E}-09$
PFIT2LS	3	238	$2.01\text{E}-08$	255	129	$1.00\text{E}-11$
PFIT3LS	3	245	$2.41\text{E}-24$	311	156	$1.18\text{E}-14$
PFIT4LS	3	419	$2.56\text{E}-13$	473	283	$2.52\text{E}-11$
POWELLSG	4	15	$4.43\text{E}-09$	16	16	$5.64\text{E}-09$
QUARTC	100	24	$2.31\text{E}-08$	27	27	$1.11\text{E}-07$
ROSENBR	2	40	$6.25\text{E}-16$	27	24	$5.93\text{E}-15$
S308	2	11	$7.73\text{E}-01$	8	8	$7.73\text{E}-01$
SBRYBND	100	30	$1.24\text{E}-13$	17	13	$3.89\text{E}-14$

Table A.1: Comparison among solution methods

		Adaptive RNM		Extended RNM		
Name	n	N_f, N_L	f	N_f	N_L	f
SCHMVETT	100	4	$-2.94\text{E}+02$	5	5	$-2.94\text{E}+02$
SINEVAL	2	123	$3.41\text{E}-15$	61	48	$2.56\text{E}-12$
SINQUAD	100	16	$-4.01\text{E}+03$	15	13	$-4.01\text{E}+03$
SISSER	2	12	$1.14\text{E}-08$	13	13	$5.90\text{E}-09$
SNAIL	2	236	$4.84\text{E}-13$	126	126	$5.58\text{E}-18$
SPARSINE	100	6	$9.52\text{E}-22$	7	7	$8.59\text{E}-21$
SPARSQUR	100	16	$7.66\text{E}-09$	17	17	$5.50\text{E}-09$
SPMSRTLS	100	10	$4.05\text{E}-11$	11	11	$1.30\text{E}-16$
SROSENBR	100	7	$7.74\text{E}-15$	14	14	$2.16\text{E}-17$
STRATEC	10	26	$2.21\text{E}+03$	38	36	$2.21\text{E}+03$
TESTQUAD	1000	4	$2.53\text{E}-20$	7	7	$4.31\text{E}-14$
TOINTGOR	50	5	$1.37\text{E}+03$	11	11	$1.37\text{E}+03$
TOINTGSS	100	5	$1.01\text{E}+01$	8	8	$1.01\text{E}+01$
TOINTPSP	50	41	$2.26\text{E}+02$	56	22	$2.26\text{E}+02$
TOINTQOR	50	4	$1.18\text{E}+03$	7	7	$1.18\text{E}+03$
TQUARTIC	100	13	$1.12\text{E}-24$	21	20	$1.73\text{E}-15$
TRIDIA	100	3	$1.60\text{E}-11$	5	5	$1.59\text{E}-14$
VARDIM	200	29	$2.53\text{E}-24$	29	29	$2.33\text{E}-25$
VAREIGVL	50	25	$2.01\text{E}-11$	12	12	$2.16\text{E}-09$
VIBRBEAM	8	44	$1.56\text{E}-01$	63	54	$1.56\text{E}-01$
WATSON	12	9	$6.60\text{E}-12$	11	11	$7.77\text{E}-09$
WOODS	4	67	$2.38\text{E}-14$	48	46	$4.09\text{E}-16$
YFITU	3	62	$6.67\text{E}-13$	221	217	$6.29\text{E}-09$
ZANGWIL2	2	4	$-1.82\text{E}+01$	5	5	$-1.82\text{E}+01$

Bibliography

[1] L. Armijo, *Minimization of functions having continuous partial derivatives*, Pacific Journal of Mathematics, 16 (1966), pp. 1–3.

[2] C. Ashcraft, R. G. Grimes, and J. G. Lewis, *Accurate symmetric indefinite linear equation solvers*, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 513–561.

[3] A. Beck and M. Teboulle, *Mirror descent and nonlinear projected subgradient methods for convex optimization*, Operations Research Letters, 31 (2003), pp. 167–175.

[4] ——, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[5] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, New York, 1995.

[6] C. Cartis, N. I. M. Gould, and P. L. Toint, *Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity*, Mathematical Programming, to appear.

[7] ——, *An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity*, IMA Journal of Numerical Analysis, to appear.

[8] ——, *Evaluation complexity of adaptive cubic regularization methods for convex unconstrained optimization*, Optimization Methods and Software, to appear.

[9] ——, *On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems*, SIAM Journal on Optimization, 20 (2010), pp. 2833–2852.

[10] ——, *Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results*, Mathematical Programming, 127 (2011), pp. 245–295.

[11] ——, *On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming*, SIAM Journal on Optimization, 21 (2011), pp. 1721–1739.

[12] ——, *Complexity bounds for second-order optimality in unconstrained optimization*, Journal of Complexity, 28 (2012), pp. 93–108.

[13] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer-Burmeister NCP-function*, Mathematical Programming, 88 (2000), pp. 211–216.

[14] J.-S. Chen, *The semismooth-related properties of a merit function and a descent method for the nonlinear complementarity problem*, Journal of Global Optimization, 36 (2006), pp. 565–580.

[15] S. H. Cheng and N. J. Higham, *A modified Cholesky algorithm based on a symmetric indefinite factorization*, SIAM Journal on Matrix Analysis and Applications, 19 (1998), pp. 1097–1110.

[16] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1983.

[17] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, 2000.

[18] R. W. Cottle, *Nonlinear Programs with Positively Bounded Jacobians*, PhD thesis, University of California, Berkeley, 1964.

[19] J. P. Crouzeix, *On second order conditions for quasiconvexity*, Mathematical Programming, 18 (1980), pp. 349–352.

[20] H. Dan, N. Yamashita, and M. Fukushima, *Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions*, Optimization Methods and Software, 17 (2002), pp. 605–626.

[21] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Mathematical Programming, 91 (2002), pp. 201–213.

[22] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Mathematical Programming, 76 (1997), pp. 493–512.

[23] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[24] F. Facchinei and J. Soares, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM Journal on Optimization, 7 (1997), pp. 225–247.

[25] M. C. Ferris and J.-S. Pang, *Engineering and economic applications of complementarity problems*, SIAM Review, 39 (1997), pp. 669–713.

[26] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[27] D. C. Gazis, *Traffic Theory*, Kluwer Academic Publishers, Boston, 2002.

[28] N. I. M. Gould, D. Orban, and P. L. Toint, *CUTEr (and SifDec), a Constrained and Unconstrained Testing Environment, revisited*, ACM Transactions on Mathematical Software, 29 (2003), pp. 373–394.

[29] S. Gratton, A. Sartenaer, and P. L. Toint, *Recursive trust-region methods for multiscale nonlinear optimization*, SIAM Journal on Optimization, 19 (2008), pp. 414–444.

[30] W. W. Hogan, *Point-to-set maps in mathematical programming*, SIAM Review, 15 (1973), pp. 591–603.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[32] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM Journal on Control and Optimization, 35 (1997), pp. 178–193.

[33] H. Jiang and D. Ralph, *Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, 1999, pp. 181–209.

[34] L. V. Kantorovich, *On Newton's method*, Trudy Matematicheskogo Instituta imeni V. A. Steklova, 28 (1949), pp. 104–144.

[35] C. Kanzow and M. Fukushima, *Equivalence of the generalized complementarity problem to differentiable unconstrained minimization*, Journal of Optimization Theory and Applications, 90 (1996), pp. 581–603.

[36] C. Kanzow, N. Yamashita, and M. Fukushima, *Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints*, Journal of Computational and Applied Mathematics, 173 (2005), pp. 321–343.

[37] K. Levenberg, *A method for the solution of certain non-linear problems in least squares*, Quarterly of Applied Mathematics, 2 (1944), pp. 164–168.

[38] D. H. Li, M. Fukushima, L. Qi, and N.Yamashita, *Regularized Newton methods for convex minimization problems with singular solutions*, Computational Optimization and Applications, 28 (2004), pp. 131–147.

[39] Y. J. Li and D. H. Li, *Truncated regularized Newton method for convex minimizations*, Computational Optimization and Applications, 43 (2009), pp. 119–131.

[40] T. D. Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Mathematical Programming, 75 (1996), pp. 407–439.

[41] C. Ma and J. Tang, *The quadratic convergence of a smoothing Levenberg-Marquardt method for nonlinear complementarity problem*, Applied Mathematics and Computation, 197 (2008), pp. 566–581.

[42] C. Ma, J. Tang, and X. Chen, *A globally convergent Levenberg-Marquardt method for solving nonlinear complementarity problem*, Applied Mathematics and Computation, 192 (2007), pp. 370–381.

[43] D. W. Marquardt, *An algorithm for least squares estimation of non-linear parameters*, SIAM Journal on Applied Mathematics, 11 (1963), pp. 431–441.

[44] J. J. Moré, *Classes of functions and feasibility conditions in nonlinear complementarity problems*, Mathematical Programming, 6 (1974), pp. 327–338.

[45] ——, *The Levenberg-Marquardt algorithm: Implementation and theory*, Numerical Analysis, 630 (1978), pp. 105–116.

[46] D. D. Morrison, *Methods for nonlinear least squares problems and convergence proofs*, in Proceedings of the Seminar on Tracking Programs and Orbit Determination, 1960, pp. 1–9.

[47] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.

[48] A. Nemirovski and D. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley, New York, 1983.

[49] Y. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Soviet Mathematics Doklady, 27 (1983), pp. 372–376.

[50] ——, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, Dordrecht, 2004.

[51] ——, *Modified Gauss-Newton scheme with worst-case guarantees for global performance*, Optimization Methods and Software, 22 (2007), pp. 469–483.

[52] ——, *Accelerating the cubic regularization of Newton's method on convex problems*, Mathematical Programming, 112 (2008), pp. 159–181.

[53] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Mathematical Programming, 108 (2006), pp. 177–205.

[54] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[55] M. R. Osborne, *Nonlinear least squares – the Levenberg algorithm revisited*, Journal of the Australian Mathematical Society, 19 (1976), pp. 343–357.

[56] J. S. Pang, *Error bounds in mathematical programming*, Mathematical Programming, 79 (1997), pp. 299–332.

[57] R. A. Polyak, *Regularized Newton method for unconstrained convex optimization*, Mathematical Programming, 120 (2009), pp. 125–145.

[58] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.

[59] P. K. Sabramanian, *Gauss-Newton methods for the complementarity problem*, Journal of Optimization Theory and Applications, 77 (1993), pp. 467–482.

[60] C. Shen, X. Chen, and Y. Liang, *A regularized Newton method for degenerate unconstrained optimization problems*, Optimization Letters, to appear.

[61] P. Tseng, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Mathematical Programming, 125 (2010), pp. 263–295.

[62] K. Ueda and N. Yamashita, *A regularized Newton method without line search for unconstrained optimization*, Technical Report, Kyoto University, 2009.

[63] ——, *Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization*, Applied Mathematics and Optimization, 62 (2010), pp. 27–46.

[64] ——, *On a global complexity bound of the Levenberg-Marquardt method*, Journal of Optimization Theory and Applications, 147 (2010), pp. 443–453.

[65] ——, *Global complexity bound analysis of the Levenberg-Marquardt method for nonsmooth equations and its application to the nonlinear complementarity problem*, Journal of Optimization Theory and Applications, 152 (2012), pp. 450–467.

[66] D. Winfield, *Function and Functional Optimization by Interpolation in Data Tables*, PhD thesis, Harvard University, Cambridge, 1964.

[67] S. J. Wright and J. N. Holt, *An inexact Levenberg-Marquardt method for large sparse nonlinear least squares*, Journal of the Australian Mathematical Society, 26 (1985), pp. 387–403.

[68] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg-Marquardt method*, Computing. Supplementum(Wien), 15 (2001), pp. 227–238.

[69] J. Zhang and X. Zhang, *A smoothing Levenberg-Marquardt method for NCP*, Applied Mathematics and Computation, 178 (2006), pp. 212–228.